Efficiency gains, bounds, and risk in finance

Sarisoy, Cisil

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Sarisoy, C. (2015). *Efficiency gains, bounds, and risk in finance*. [Doctoral Thesis, Tilburg University]. CentER, Center for Economic Research.

Efficiency Gains, Bounds, and Risk in Finance

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op donderdag 17 december 2015 om 10.15 uur door

ÇİSİL SARISOY

geboren op 9 juli 1986 te Zonguldak, Turkije.

Promotor: Prof. Dr. Bas Werker

Copromotor: Dr. Peter de Goeij

Overige leden van de promotiecommissie:

Prof. Dr. Frank de Jong

Prof. Dr. Joost Driessen

Prof. Dr. Eric Renault

Prof. Dr. Jules van Binsbergen

Acknowledgements

A great many people contributed to the completion of this Ph.D. thesis in different ways over the last years. It has been an intellectually challenging and enjoyable period of my life and I would like to thank all those people involved during this journey.

First and foremost, I would like to express my sincerest gratitude to my supervisor, Bas J.M. Werker, who has played a vital role in my learning process since the research master program at Tilburg University. He always provided valuable feedback on my work and his openness and interest in different topics played an important role in shaping a wide range of research interests for me. It was very important and crucial for my academic development that Bas was always open to discuss my questions, and helped me how and where to find answers. Moreover, he gave me great feedback on presenting and writing technical research in a structured and understandable manner for diverse audiences. His attention to details have set a high standard which I try to live up to during my own academic life. I am also thankful to him for always being such a positive person and believing in my success. Last but not least, besides research, he was very kind and supportive during a particularly difficult period of my life. I am especially grateful for that.

I would also like to express my appreciation and thanks to Peter de Goeij, my copromotor, for his valuable advice, and support; in particular for encouraging me to think about economic intuition which has also shaped my view about evaluating and conducting research. Moreover, I would like to thank Eric Renault. I met Eric Renault during my first year as a Ph.D. student and I am very grateful to have the honour to be working with him till this day and share his insights in the field of volatility research and Econometrics more generally.

I am grateful to the members of my dissertation committee, Frank de Jong, Joost Driessen, Eric Renault, and Jules van Binsbergen, for their brilliant comments,

suggestions and advice. It is an honour for me to have this group as my committee members.

I would like to thank to the Department of Finance, Tilburg University, for giving me the opportunity to present my work at various conferences, for the valuable feedback on my work and also for providing a stimulating research environment. In particular, I would like to mention my thanks to Frans de Roon and Lieven Baele for many insightful discussions.

Part of this research has been conducted during my time as a visiting Ph.D. candidate at the Kellogg School of Management, Northwestern University. I'm indebted to Torben G. Andersen, my host, for making this visit possible, and I am looking forward to joint work in the future. I would also like to thank the faculty members of the Finance Department at Kellogg for the many discussions about research and their hospitality during my visit. I am grateful to the Kellogg School of Management for the facilities that they provided during my visit, and NWO and CentER for financial support.

I would like to also thank all the people that made the years of my Ph.D. in Tilburg so enjoyable. I am thankful to my housemates, Regina, Sara and Yanting for creating a family–like environment at home in Tilburg. My Ph.D. cohort in Tilburg has been a source of good friendships as well as advice and collaboration about work. I would like to mention Anton, Bernardus, Larissa, Paola, Jiehui and Ran. I very much enjoyed our talks, our activities, which made a quiet city like Tilburg lively and fun - thank you all. I had two great office mates; Cara and Elena, thanks for being so supportive, helpful and fun office mates. Moreover, I would like to thank my friends scattered around the world for their advice, support and being there whenever I needed a friend. Special thanks go to Bilge, Burcu, Derya, Ece, Paulina and Tuğce.

Finally, I would like to thank my family in Turkish. Sevgili kardesim, anne, baba, Nural hala ve Bora: boylesine sevgi dolu bir aile oldugunuz, tum guzel ve

zor zamanlarda yanimda oldugunuz, ve egitim hayatim boyunca sagladiginiz tum imkanlar ve desteginiz icin sizlere cok tesekkur ederim. Asli abla, Ahu abla, Fikriye teyze, Nursen hala, Onur ve Ozan: tum egitim hayatim suresince ilginiz, sevginiz, onerileriniz ve desteginiz icin cok tesekkur ederim. Basarilarimda sizlerin verdiginiz motivasyonun cok buyuk bir etkisi var. Sabine: thank you for your support during the years of my Ph.D.. Lastly, I am indebted to my beloved husband Tristan for his great support, patience and being there as a sparring partner. I very much enjoy our stimulating discussions about research which open up new perspectives.

Contents

Introduction

Financial theory often inherently relies on the assumption that the investor knows the true parameters of a model capturing the asset price dynamics of interest. However, real life applications require the estimation of the unknown parameters. When applying models to data, the user has to understand the connection between the theoretical formulation of a model and its empirical solution in the data. The applicability of a model and, thus, its value are partly determined by how well the parameters can be estimated. Consequently, estimation risk is inherently unavoidable whenever models are to be run on the data and it stands as an important issue in the field of financial economics.

The naturally arising question then concerns the minimization of this risk. One common approach is to seek for smallest variances i.e., in statistical jargon, asymptotic efficiency. To accomplish such a task, researchers analyze the asymptotic distributions of the parameters, in particular the asymptotic variances, and analyze the performance of estimates based on their asymptotic variances. From a practical perspective, the aim is to understand which estimators have smaller asymptotic variances, in other words which estimators are (asymptotically) more efficient. This approach is motivated by the desire to obtain smaller standard errors and thus smaller confidence intervals.

The first chapter of this thesis, essentially equivalent to the working paper under the same title co–authored with Peter de Goeij and Bas J.M. Werker, concerns the estimation of expected returns, which is, perhaps, one of the longest standing questions in finance. Expected returns are not only interesting in the sense of single quantities for individual assets but they are also crucial inputs for theoretical formulations of problems in various subfields of finance. From a corporate finance perspective, they are key inputs for calculating cost of capital as well as for the valuation of cash flows. We require estimates of expected returns to obtain the required rate of return or to discount the payoffs or cash flows of an asset. From an asset pricing perspective, the most prominent presence of expected returns is in portfolio allocation decisions.

Asset pricing theory provides a theoretical foundation regarding the cross-section of expected returns based on equilibrium models, partial equilibrium models and reduced form specifications such as multifactor models. These models motivate certain risks that explain the cross section of expected returns on assets. Potentially, expected returns can be estimated by imposing equilibrium restrictions of these models. But is this approach useful? Does imposing asset pricing models bring any advantages in estimating expected returns over standard methods?

The main objective of the first chapter is to understand the benefits of asset pricing models in estimating expected returns. In particular, the chapter provides an analysis of the efficiency gains by imposing the restrictions of asset pricing models, with a particular focus on linear factor models. One might ask whether it is necessary to obtain an(other) estimate of expected returns which achieves efficiency gains? This is indeed the case and the literature needs additional guidance on this issue because the traditional estimate at hand, i.e. historical averages, has been shown to be a very noisy estimate. This translates into the need for a very large, in practice mostly infeasible, samples of data in order to gain a bit from precision. Therefore, having a more efficient estimator of expected returns would be a good help not only for academics but also for practitioners for understanding the solutions of the theoretical formulations in the data.

The first chapter provides not only the asymptotic distributions of the expected return estimators based on factor models but also, closed form asymptotic expressions for analyzing the efficiency gains over historical averages. The decision maker believing in an asset pricing model can plug in the parameter estimates from her asset pricing model and calculate the efficiency gains of expected return estimates based on the factor model over the historical averages. In the standard Fama-French three factor model (MKT, SMB, HML) setting with 25 FF-portfolios, the first chapter of the thesis documents that the efficiency gains are 36% on average across these 25 portfolios, even increasing up to 50% for certain portfolios. For real life applications, this translates into the benefit of using only half the data with factor model based estimates to obtain the same precision as with historical averages.

What are the economic implications? The second part of the first chapter of this thesis analyzes the implications of using factor model based estimates of expected returns for portfolio allocation problems in Markowitz's (1952) setting. The literature documents that the imprecise estimates of expected returns, via historical averages, leads to an economically significant deterioration of the out-of-sample performance of the optimal portfolios.

In the far end, this has led to simply abandoning the application of theoretically optimal decisions and using naive techniques such as the 1/N portfolio or the global minimum portfolio as these are not subject to estimation risk from estimating expected returns. Although in the setting of Markowitz (1952), optimal portfolios are supposed to achieve the best performance, in practice they turn out to perform worse due to imprecise estimates of expected returns and of the covariance matrix. The literature provides some solutions on improving the covariance matrix estimates, however, researchers have been much less active on providing guidance of expected return estimates, which has been documented as the driving source of the problem in such a context (De Miguel et al. (2009)). The first chapter documents that out-of-sample Sharpe ratios of the optimal portfolios almost doubles if factor model based estimates are used instead of historical averages. It is also interesting to see in the simulation exercise that these figures come close to their theoretical values and outperform the global minimum variance portfolios as well as the 1/N portfolios. The first chapter of the thesis provides a base for such a result by providing the asymptotic efficiency gains of factor model based expected return estimates and aims to provide guidance for academics as well as practitioners.

The second chapter of the thesis, essentially equivalent to the manuscript under the same title co–authored with Eric Renault and Bas J.M. Werker (accepted at Econometric Theory), is situated in a fast growing area of research: the study of high frequency data which is, perhaps, likely to become an overarching theme in the field of finance ranging from risk management over derivative pricing to portfolio management from both empirical and theoretical perspective. As the chapter is quite specialized, I will give an overview of the literature and provide an intuition of it is important first.

The key ingredients underlying any empirical application making use of high frequency data are functionals of variation of "realized" asset price paths. An important advantage of such approach is that these estimators are model–free in the sense that they do not require estimation of parametric and thus potentially restrictive models. The increasing availability of high frequency data at the intraday level has spurred the development and analysis of high frequency based estimators for integrated volatility and their asymptotic properties starting with Andersen and Bollerslev (1998), Comte and Renault (1998), and Barndorff-Nielsen and Shephard (2001, 2002a,b) among others. The chapter focuses in particular on the efficient estimation of integrated smooth transformations of volatility and related processes.

Estimators of integrated variance, quadratic variation, and related quantities have shown to be of crucial importance for many standard applications in finance. The most straightforward improvements are gains in terms of volatility forecasting as shown in a series of papers, i.e. Andersen et al. (2003), Andersen et al. (2004), Andersen et al. (2007), Ghysels et al. (2006), among many others. The benefits of using high frequency based volatility estimators are also mirrored in superior density forecasts over EGARCH models based on daily data (Maheu and McCurdy, 2011) and in Value-at-Risk estimation (Andersen et al., 2003). From an asset allocation perspective, incorporating such estimators leads to a) significant performance gains with regard to the global minimum variance portfolio compared to methods based on daily data (Hautsch et al., 2015) and b) utility gains for a mean variance investor by using realized variance calculated at different sampling frequencies (Bandi and Russell, 2006).

As the data has become richer and researchers are equipped with empirical and theoretical insights, the research agenda has become more intricate: several model–free estimators, not only for integrated variance but also for integrated power variances and other smooth transformations, have been suggested. These are interesting quantities for hypothesis testing when one seeks to infer and analyze the precision of integrated volatility estimators in terms of their confidence intervals. In particular, the asymptotic distribution of realized variance depends on the unknown integrated quarticity. As higher power variations are naturally more noisy objects than lower power variations such as realized variance, this calls for the need to estimate them with high precision (see, e.g., Jacod and Rosenbaum (2013) and Andersen et al. (2014) for an overview regarding the estimation of integrated quarticity). The usefulness of these higher order quantities also extends to various tests for detecting discontinuities in realized asset price paths, which has produced mixed results, see the literature survey in Christensen et al. (2014). Integrated quarticity also serves as tuning parameter in the bandwidth selection of microstructure noise robust realized kernel estimation of integrated quarticity (Barndorff–Nielsen et al., 2009). Considering other smooth transformations, Realized Laplace transforms are provided to make inferences about spot volatility dynamics; see Todorov and Tauchen (2012), Todorov, Tauchen, and Grynkiv (2011).

The literature provides more efficient estimators for integrated smooth transformations of instantaneous variances not only in simpler settings such as equally spaced observations and exclusion of jumps, which makes it technically easier and more convenient to understand gains, but also in more complicated and more realistic settings such as random but unequally spaced observation times, which makes the technique conceptually difficult.

The open question in the literature has been to understand the concept of optimality in non-parametric settings, which can be understood as a bound on the asymptotic variances of the estimators. In simpler terms, a natural question would be how to understand if there is a better estimator than the ones already provided in the literature given a very general data generating processes commonly used in the literature. Should we still search for new estimators which are maybe more efficient than the other?

The second chapter of this thesis provides efficiency bounds, i.e. lower bounds on the precision of the regular estimators for integrated variances and also for other smooth transformations of instantaneous variances such as integrated power variances, and Laplace transforms. Jacod and Rosenbaum (2013) mention that "... even for the simpler problem of estimating integrated volatility, the concept of efficiency in the general non-parametric or semi-parametric setting is not well established so far." The second chapter of thesis provides these bounds in general settings, where the times may be unequally sampled or there may be jumps in the volatility processes.

The natural next step would be to analyze existing estimators of the literature if they are optimal. It follows from our results that, for integrated variance, Realized Variance is a non–parametrically efficient estimator for integrated variance under both regular and irregular sampling schemes. Given our results, it turns out that in case of irregularly spaced observations, the literature has not yet been able to provide an efficient estimator for other integrated smooth transformations and integrated power variances other than integrated variance. A nearly efficient kernel based estimator is provided in the second part of the second chapter to narrow this gap.

The third chapter of the thesis is on a slightly different topic than the other two. It focuses on a topic that has been very active recently: monetary policy and risk in financial markets. The aim of U.S. monetary policy is defined in terms of macroeconomic aggregates, in particular price stability, maximum employment and output. The policy maker, here the Federal Reserve, takes actions through instruments which are at best indirectly geared towards achieving those goals. Bernanke and Kuttner (2005) further states that "by affecting asset prices and returns, policy makers try to modify economic behaviour in ways that will help to achieve their ultimate objectives." The naturally arising challenge is to resolve the form of connecting links, if any, between these three variables 1.) policy making decisions, 2.) asset prices, and 3.) economic activity.

There has been a considerable interest in understanding the time series relations between the Fed Funds target rate announcements and asset returns in fixed income, foreign exchange, and aggregate equity markets; see, among others Kuttner (2001), Andersen et al. (2003), Rigobon and Sack (2004), Bernanke and Kuttner (2005), and Gürkaynak et al. (2005). However, surprisingly, there has been relatively less attention to understand the links between monetary policy shocks and the cross–section of expected returns.

Recent literature documents large average excess returns of the U.S. equity market on announcement days of FOMC interest rate decisions. Savor and Wilson (2013) documents that on the days of employment, inflation and FOMC announcements, equity market experiences significantly larger average excess returns compared to other days. In particular, they find that over 60% of the cumulative annual equity risk premium is earned on these announcement days. However, it seems likely that these results are driven by the large magnitude of the equity market average excess returns on FOMC announcement days; see Table 3.1 of the third chapter. Moreover, Lucca and Moench (2015) documents that 80% of the realized stock excess returns have been earned in the 24 hour pre-announcement period. A closer look at their results in their Table II reveals that a large proportion of these returns are earned within the announcement days.

Given the substantial amount of average excess returns exhibited by US equity market on FOMC days, a naturally arising question would regard if these returns represent any compensation for being exposed to monetary policy risk. Chapter 3 of the thesis analyzes this question and seeks to provide answers if monetary policy risks are priced in the cross–section of stocks.

As it is unlikely that stock prices respond to anticipated information about policy actions, monetary policy shocks are defined as the "surprise" component in target rate changes. Moreover, in order to estimate exposures of individual stock returns to factors as precisely as possible, the intraday data is employed. In particular, intraday event windows around the FOMC press releases are used to measure the response of individual stock prices to monetary policy shocks. A proof is provided in Appendix 3.A.1, detailing why such approach would lead to precision gains in estimation of exposures.

The results of the chapter show that shocks to monetary policy carry a statistically significant negative price of risk. This translates to stocks which are positively (negatively) exposed to monetary policy shocks earning lower (higher) average returns, all else being equal. Moreover, Chapter 3 also provides an analysis of the prices of risk at the intraday level, in particular for three distinct intraday observation windows on announcement days: pre–announcement window, announcement window and post–announcement window. The results show that most of the monetary policy risk premium and the market risk premium are earned during the pre–announcement windows. This result is in line with the findings of Lucca and Moench (2015), who illustrates large market excess returns prior to the announcements.

Chapter 1

Linear Factor Models and the Estimation of Expected Returns

Abstract

Estimating expected returns on individual assets or portfolios is one of the most fundamental problems of finance research. The standard approach, using historical averages, produces noisy estimates. Linear factor models of asset pricing imply a linear relationship between expected returns and exposures to one or more sources of risk. We show that exploiting this linear relationship leads to statistical gains of 36% in standard deviations when estimating expected returns over historical averages. If the factor model is misspecified in the sense of an omitted factor, we show that factor model–based estimates may be inconsistent. However, we show that adding an alpha to the model capturing mispricing only leads to consistent estimators in case of traded factors. Moreover, our simulation experiment shows that using factor–model based estimates of expected returns significantly improves the out–of–sample performance of the optimal portfolios.

1.1 Introduction

One of the key problems of finance studies is the estimation of risk premiums, that is expected excess returns, on individual securities or portfolios. The standard approach, which has been favoured by researchers, investors and analysts, is to use historical averages. However, it is also known that these estimates are generally very noisy. Even using daily data does not help much, if at all. One needs very long samples for accurate estimates, which often are unavailable.

The asset pricing literature provides a wide variety of linear factor models motivating certain risks that explain the cross section of expected returns on assets. Examples include Sharpe (1964)'s CAPM, Merton (1973)'s ICAPM, Breeden (1979)'s CCAPM, the arbitrage pricing theory of Ross (1976a,b), Lettau and Ludvigson's (2001) conditional CCAPM among many others. These models all imply that expected returns of assets are linear in their exposures to the risk factors. The coefficients in this linear relationship are the prices of the risk factors. The literature on factor models mainly concentrates on determining these prices of risk and evaluating the ability of the models in explaining the cross section of expected returns on assets.

In this study, the focus is different: we assess the precision gains in the estimation of the expected (excess) returns on an individual asset and on portfolios, i.e., the product of exposures (β) and risk prices (λ), vis–à–vis the historical averages approach. As mentioned by Black (1993), theory can help to improve the estimates of expected returns. We show when exploiting the linear relationship implied by linear factor models indeed leads to more precise estimates of expected returns over historical averages.

Estimating expected returns using factor models is not a new idea and was, to our knowledge, first suggested by Jorion (1991). In his empirical analysis, he compares CAPM—based estimators with classical sample averages of past returns finding the former outperforming the latter in estimating expected stock returns for his data. Our paper complements his work by providing the first detailed asymptotic efficiency analysis for both estimators, and evaluating the implications of omitted factors on the estimation of expected (excess) returns.

First, we investigate the issue of accurate estimation of risk–premiums, i.e. expected excess returns on individual assets and portfolios by providing a detailed (asymptotic) analysis of risk–premium estimators based on factor models. Comparing the limiting covariance matrices of factor–based risk–premium estimators with those of the historical averages estimator, we find sizeable efficiency gains from imposing the factor structure, see Corollaries (4.1-4.2). In an empirical analysis, for instance when estimating risk–premiums on 25 size and book–to–market Fama–French portfolios, we document large gains in standard deviations of 36% on average.

Secondly, we consider the issue of estimating risk premiums in the ubiquitous situation where one may face omitted factors in the specification of the linear factor model. After the capital asset pricing model (CAPM) has been substantially criticized, researchers have come up with new risk factors to help explaining the cross section of expected returns, e.g., Fama and French (1993), Lettau and Ludvigson (2001), Lustig and Van Nieuwerburgh (2005), Li, Vassalau and Xing (2006), Santos and Veronesi (2006). While it is doubtful that "the correct" factors have been found, the literature points to the existence of missing factors. We show that when a model is misspecified, in the sense that a relevant pricing factor is omitted, standard methods will generally not even provide consistent estimates of risk premiums on the individual assets or portfolios (see Theorem 1.3). However, we show that adding an alpha capturing the misspecification leads to a consistent estimator only in case of traded factors, but there is no efficiency gain over historical averages. Thus, our paper documents precisely the trade-off any empirical researcher always faces: allow for misspecification and loose efficiency or run the risk of misspecification and gain efficiency.

The mean—variance framework of Markowitz (1952) is still a very popular model for portfolio allocation used in practice. However, it is also well known that the practical applications suffer from uncertainty in the parameter estimates. In particular, portfolios constructed with sample counterparts of first two moments in general have poor out of sample performance.[1] Merton (1980), followed by Chopra and Ziemba (1993), pointed out that estimation error in asset return means is more severe than errors in covariance estimates. Moreover, imprecision in estimates of the mean has a much larger impact on portfolio weights compared to the imprecision in covariance estimates (DeMiguel et al., 2009)). The mean—variance portfolio weights could also be constructed with factor–based risk–premium estimates instead of the "naive" estimates (historical averages). Accordingly, we investigate if it is possible to achieve performance gains based on the higher precision of factor–based risk–premium estimates. In particular, we analyze the out—of—sample performances of tangency portfolios based on various risk–premium estimators in a simulation study. Our results document that the average out–of–sample Sharpe ratio of the tangency portfolio increases strikingly if the portfolio weights are constructed with factor–based risk–premium estimates rather than the naive estimates. Moreover, out–of–sample Sharpe ratios of the factor–based tangency portfolios is more precise than the tangency portfolios based on historical averages. Our simulation results also document that these portfolios, in contrast to the tangency portfolios based on historical averages, perform considerably better than the global minimum variance portfolio.

The rest of the paper is organized as follows. Section 1.2 introduces our set–up and presents the linear factor model with the assumptions that form the basis of our statistical analysis. Next, we introduce factor–mimicking portfolios and clarify the link between the expected return obtained with non–traded factors and with factor–mimicking portfolios. Section 1.3 discusses in detail the estimators we consider. In particular, we recall the different sets of moment conditions for various cases such

[1] See, for example, Frost and Savarino (1986, 1988), Michaud (1989), Jobson and Korkie (1980, 1981), Best and Grauer (1991), and Litterman (2003).

as all factors being traded and factor–mimicking portfolios. Section 1.4 derives the asymptotic properties of these induced GMM estimators. In particular, we derive the efficiency gains over and above the risk–premium estimator based on historical averages. Section 5 adresses the question of using misspecified factor pricing models. Section 1.6 documents the simulation analysis for portfolio optimization, while Section 3.5 concludes. All proofs are gathered in the appendix.

1.2 Model

It is well known that in the absence of arbitrage, there exists a stochastic discount factor M such that for any traded asset $i = 1, 2, \ldots, N$ with excess return R_i^e

$$\mathrm{E}\left[MR_i^e\right] = 0. \tag{1.2.1}$$

Linear factor models additionally specify $M = a + b'F$, where $F = (F_1, ..., F_K)'$ is a vector of K factors (see, e.g., Cochrane (2001), p.69). Note that (1.2.1) can be written in matrix notation using the vector of excess returns $R^e = (R_1^e, ..., R_N^e)'$. Throughout we impose the following.

Assumption 1.1. *The N–vector of excess asset returns R^e and the K–vector of factors F with K<N satisfy the following conditions:*

1. *The covariance matrix of excess returns $\Sigma_{R^eR^e}$ has full rank N,*
2. *The covariance matrix of factors Σ_{FF} has full rank K,*
3. *The covariance matrix between excess returns and factors $Cov\left[R^e, F'\right]$ has full rank K.*

Given the linear factor model and Assumption 1.1, it is classical to show

$$\mathrm{E}\left[R^e\right] = \beta\lambda, \tag{1.2.2}$$

where

$$\beta = \mathrm{Cov}\left[R^e, F'\right]\Sigma_{FF}^{-1}, \tag{1.2.3}$$

$$\lambda = -\frac{1}{\mathrm{E}\left[M\right]}\Sigma_{FF}b. \tag{1.2.4}$$

Thus, (1.2.2) specifies a linear relationship between risk premiums, $\mathrm{E}\left[R^e\right]$, and the exposures β of the assets to the risk factors, F, with prices λ.

In empirical work, we need to make assumptions about the time–series behavior of consecutive returns and factors. In this paper, we focus on the simplest, and most used, setting where returns are i.i.d. over time. Express the excess asset returns

$$R_t^e = \alpha + \beta F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1.2.5}$$

where α is an N–vector of constants, ε_t is an N–vector of idiosyncratic errors and T is the number of time–series observations. We then, additionally, impose the following.

Assumption 1.2. *The disturbance ε_t and the factors F_t, are independently and identically distributed over time with*

$$E\left[\varepsilon_t | F_t\right] = 0, \tag{1.2.6}$$

$$Var\left[\varepsilon_t | F_t\right] = \Sigma_{\varepsilon\varepsilon}, \tag{1.2.7}$$

where $\Sigma_{\varepsilon\varepsilon}$ has full rank.

1.2.1 Factor–Mimicking Portfolios

A large number of studies in the asset pricing literature suggest "macroeconomic" factors that capture systematic risk. Examples include the C-CAPM of Breeden (1979), the I-CAPM of Merton (1973) and the conditional C-CAPM of Lettau and Ludvigson (2001). In order to assess the validity of macroeconomic risk factors being priced or not, it has been suggested to refer to alternative formulations of such factor models replacing the factors by their projections on the linear span of the returns. This is commonly referred to as factor mimicking portfolios and early references go back to Huberman et al. (1987) and also compare, e.g., Fama (1998) and Lamont (2001). We analyze, in this paper, the role of such formulations on the estimation of risk premiums and we show, in Section 1.4, that there are efficiency gains from the information in mimicking portfolios in estimating risk premiums.

We project the factors F_t onto the space of excess asset returns, augmented with a constant. In particular, given Assumption 1.1, there exists a K–vector Φ_0 and a $K \times N$ matrix Φ of constants and a K–vector of random variables u_t satisfying

$$F_t = \Phi_0 + \Phi R_t^e + u_t, \tag{1.2.8}$$

$$\mathrm{E}\left[u_t\right] = 0_{K\times 1}, \tag{1.2.9}$$

$$\mathrm{E}\left[u_t R_t^{e\prime}\right] = 0_{K\times N}, \tag{1.2.10}$$

and we define the factor–mimicking portfolios by

$$F_t^m = \Phi R_t^e. \tag{1.2.11}$$

We, then, obtain an alternative formulation of the linear factor model by replacing the original factors with factor–mimicking portfolios[2]

$$R_t^e = \alpha^m + \beta^m F_t^m + \varepsilon_t^m, \quad t = 1, 2, \ldots, T. \tag{1.2.12}$$

Recall that using the projection results, Φ and β are related by

$$\Phi = \Sigma_{FF}\beta'\Sigma_{R^eR^e}^{-1}, \tag{1.2.13}$$

[2]Compare Huberman et al. (1987), and Kandel and Stambaugh (1995), Balduzzi and Robotti (2008), Hou and Kimmel (2010).

while β^m and β satisfy

$$\beta^m = \beta \left(\beta' \Sigma_{R^e R^e}^{-1} \beta\right)^{-1} \Sigma_{FF}^{-1}. \tag{1.2.14}$$

The following theorem recalls that, while factor loadings and prices of risk change when using factor mimicking portfolios, expected (excess) returns, their product, are not affected. For completeness we provide a proof in the appendix.

Theorem 1.1. *Under Assumptions 1.1 and 1.2, we have $\beta\lambda = \beta^m \lambda^m$, where $\lambda^m = E[F_t^m]$.*

Note that since the factor–mimicking portfolio is an excess return, asset pricing theory implies that the price of risk attached to it, λ^m, equals its expectation. This can be imposed in the estimation of expected (excess) returns and thus one may hope that the expected (excess) return estimators obtained with factor–mimicking portfolios are more efficient than the expected (excess) return estimators obtained with the non-traded factors themselves.

1.3 Estimation

As indicated in the introduction, we concentrate on Hansen's (1982) GMM estimation technique. The GMM approach is particularly useful in our paper as it avoids the use of two-step estimators and the resulting "errors-in-variables" problem when calculating limiting distributions. In addition, we immediately obtain the joint limiting distribution of estimates for β and λ which is needed as we are interested in their product.

In the following sections, we study the asymptotics of the expected (excess) return estimators by specifying different sets of moment conditions. In Section 1.3.1, we study a set of moment conditions which generally holds, i.e., both when factors are traded and when they are non-traded. In Section 1.3.2, we study the case where all factors are traded. We then incorporate the moment condition that factor prices equal expected factor values. In Section 1.3.3, we consider expected (excess) return estimates based on factor–mimicking portfolios.

1.3.1 Moment Conditions - General Case

We first provide the moment conditions for a general case, i.e., where factors may represent excess returns themselves, but not necessarily. In that case, the resulting moment conditions to estimate both factor loadings β and factor prices λ are

$$\mathrm{E}\left[h_t(\alpha,\beta,\lambda)\right] = \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \left[R_t^e - \alpha - \beta F_t\right] \\ R_t^e - \beta\lambda \end{bmatrix} = 0. \qquad (1.3.1)$$

The first moment conditions identifies α and β as the regression coefficients, while the last conditions represent the pricing restrictions. Note that there are $N \times (1+K+1)$ moment conditions although there are $N \times (1+K)+K$ parameters, which implies that the system is overidentified. Again following Cochrane (2001), we set a linear combination of the given moment conditions to zero, that is, we set $A\mathrm{E}\left[h_t(\alpha,\beta,\lambda)\right] = 0$, where

$$A = \begin{bmatrix} I_{N(1+K)} & 0_{N(1+K)\times N} \\ 0_{K\times(KN+N)} & \Theta_{K\times N} \end{bmatrix}.$$

Note that the matrix A specified above combines the last N moment conditions into K moment conditions so that the system becomes exactly identified. Following Cochrane (2001), we take $\Theta = \beta^T \Sigma_{\varepsilon\varepsilon}^{-1}$. The advantage of this particular choice is that the resulting λ estimates coincide with the GLS cross–sectional estimates.

1.3.2 Moment Conditions - Traded Factor Case

Asset pricing theory provides an additional restriction on the prices of risk when factors are traded, meaning that they are excess returns themselves. If a factor is an excess return, its price equals its expectation. For example, the price of market risk is equal to the expected market return over the risk–free rate, and the prices of size and book–to–market risks, as captured by Fama-French's SMB and HML portfolio movements, are equal to the expected SMB and HML excess returns. Note that we use the term "excess return" for any difference of gross returns, that is, not only in excess of the risk-free rate. Prices of excess returns are zero, i.e., excess returns are zero investment portfolios.

The standard two pass estimation procedure commonly found in the finance literature may not give reliable estimates of risk prices when factors are traded. Hou and Kimmel (2010) provide an interesting example to point out this issue. They generate standard two pass expected (excess) return estimates (both OLS and GLS) in the three factor Fama–French model by using 25 size and book–to–market porfolios as test assets. As shown in their Table 1, both OLS and GLS risk price estimates of the market are significantly different from the sample average of the excess market return. It is important to point out that the two pass procedure ignores the fact that the Fama–French factors are traded factors and it treats them in the same way as non–traded factors.

Consequently, when factors are traded we may use the additional moment condition that their expectation equals λ. Then, the relevant moment conditions are given by

$$\mathrm{E}\left[h_t(\alpha,\beta,\lambda)\right] = \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \left[R_t^e - \alpha - \beta F_t\right] \\ F_t^e - \lambda \end{bmatrix} = 0, \tag{1.3.2}$$

where F_t is the $K \times 1$ vector of factor (excess) returns.

In this case, estimates are obtained by an exactly identified system, i.e., number of parameters equals the number of moment conditions. Note that if the factor is traded, but we do not add the moment condition that the factor averages equal λ , then the results are just those of the non-traded case in Section 1.3.1.

Note that alternatively, we could incorporate the theoretical restriction on factor prices into the estimation by adding the factor portfolios as test assets in the linear pricing equation, $R^e - \beta\lambda$. This set of moment conditions would be similar to the general case, with the only difference being that the linear pricing restriction incorporates the factors as test assets in addition to the original set of test assets. Under this setting, the moment conditions would be given by

$$\mathrm{E}\left[h_t(\alpha, \beta, \lambda)\right] = \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \left[R_t^e - \alpha - \beta F_t\right] \\ R_t^e - \beta_{F,R}\lambda \end{bmatrix} = 0, \qquad (1.3.3)$$

where $\beta_{F,R} = \begin{bmatrix} \beta \\ I_K \end{bmatrix}$. Following the same procedure as in the general case, we specify an A matrix and set $\Theta = \beta_{F,R}^T \Sigma_{R^F R^F}^{-1}$ with $R^F = \begin{bmatrix} R_t^e \\ F_t \end{bmatrix}$. Because we find that the GMM based on (1.3.3) leads to the same asymptotic variance covariance matrices for risk premiums as the GMM based on (1.3.2), we omit the GMM based on (1.3.3) in the rest of the paper and present results for the GMM based on (1.3.2).

1.3.3 Moment Conditions - Factor–Mimicking Portfolios

Following Balduzzi and Robotti (2008), we also consider the case where risk prices are equal to expected returns o factor–mimicking portfolios. Then, the moment conditions to be used are

$$\mathrm{E}\left[h_t(\alpha^m, \beta^m, \Phi_0, \Phi, \lambda^m)\right] = \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ R_t^e \end{bmatrix} \otimes \left[F_t - \Phi_0 - \Phi R_t^e\right] \\ \begin{bmatrix} 1 \\ F_t^m \end{bmatrix} \otimes \left[R_t^e - \alpha^m - \beta^m F_t^m\right] \\ \Phi R_t^e - \lambda^m \end{bmatrix} = 0, \qquad (1.3.4)$$

with $F_t^m = \Phi R_t^e$. In this case, there are $K(1+N)+N(1+K)+K$ moment conditions and parameters, which makes the system again exactly identified.

1.4 Precision of Risk–Premium Estimators

As mentioned in the introduction, our focus is on estimating risk premiums of individual assets or portfolios. However, much of the literature on multi–factor asset pricing models has primarily focused on the issue of a factor being priced or not. Formally, this is a test on (a component of) of λ being zero or not and, accordingly, the properties of risk price estimates for λ have been studied and compared. Examples include Shanken (1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kleibergen (2009), Lewellen, Nagel and Shanken (2010), Kan and Robotti (2011), and Kan et al. (2013).

In the current paper, since our focus is on analyzing the possible efficiency gains based on linear factor models in estimating expected (excess) returns, we first derive the joint distribution of estimates for β and λ for the three GMM estimators introduced in Sections 1.3.1 to 1.3.3. Then, we derive the asymptotic distributions of the implied expected (excess) return estimators given by the product $\hat{\beta}\hat{\lambda}$. Moreover, we illustrate the empirical relevance of our asymptotic results using the Fama–French three factor model with 25 Fama–French size and book–to–market portfolios as test assets. In particular, we provide the (asymptotic) standard deviations of the various risk–premium estimators with empirically reasonable parameter values and evaluate the benefits of using linear factor models in estimating risk premiums; see Table 1.1.

Data for Empirical Results: The asset data used in this paper consists of 25 portfolios formed by Fama and French (1992, 1993), downloaded from Kenneth French‘s website. These portfolios are value–weighted and formed from the intersections of five size and five book–to–market (B/M) portfolios and they include the stocks of the New York Stock Exchange, the American Stock Exchange, and NASDAQ. For details, we refer the reader to the Fama and French (1992, 1993). The factors are the 3 factors of Fama and French (1992) (market, book–to–market and size). Our analysis is based on monthly data from January 1963 until October 2012, i.e., we have 597 observations for each Fama–French portfolio.

The following theorem provides the limiting distribution of the historical averages estimator. It's classical and provided for reference only.

Theorem 1.2. *Given that $R_1^e, R_2^e, ..., R_t^e$ is a sequence of independent and identically distributed random vectors of excess returns, we have $\sqrt{T}\left(\bar{R}^e - E[R^e]\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{R^eR^e})$.*

Note that Theorem 1.2 assumes no factor structure. We will, next, provide the asymptotic distributions of expected (excess) return estimators given the linear factor structure implied by the Asset Pricing models. Note that the joint distributions of λ and β are different for each set of moment conditions, which leads to different asymptotic distributions. Hence, we derive the asymptotic distributions of expected (excess) return estimators for the three set of moment conditions introduced in Sections 1.3.1, 1.3.2 and 1.3.3 separately.

1.4.1 Precision with General Moment Conditions

The following theorem provides the asymptotic variances of the risk–premium estimators based on the general moment conditions as in Section 1.3.1. Note that this result is valid for both traded and non-traded factors.

Theorem 1.3. *Impose Assumptions 1.1 and 1.2, and consider the moment conditions (1.3.1)*

$$E[h_t(\alpha, \beta, \lambda)] = E\begin{bmatrix} \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes [R_t^e - \alpha - \beta F_t] \\ R_t^e - \beta\lambda \end{bmatrix} = 0.$$

Then, the limiting variance of the expected (excess) return estimator $\hat{\beta}\hat{\lambda}$ is given by

$$\Sigma_{R^e R^e} - \left(1 - \lambda' \Sigma_{FF}^{-1} \lambda\right) \left(\Sigma_{\varepsilon\varepsilon} - \beta(\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \beta)^{-1} \beta'\right). \tag{1.4.1}$$

The proof is provided in the appendix. Theorem 1.3 provides the asymptotic covariance matrix of the factor–model based risk–premium estimators with the general moment conditions as in Section 1.3.1. This formula is useful mainly for two reasons. First, it can be used to compute the standard errors of these risk–premium estimates and, accordingly, the related t–statistics can be obtained. Second, it allows us to study the precision gains for estimating the risk premiums from incorporating the information about the factor model.

In case of a one–factor model and there is one–test asset, the (asymptotic) variances of both the naive risk–premium estimator and the factor–model based risk–premium estimator with (1.3.1) are the same. When more assets/portfolios are available, $N > 1$, observe that size of the asymptotic variances of risk–premium estimators depends on the magnitude of the prices of risk associated with the factor λ (per unit variance of the factor), the exposures β, and $\Sigma_{\varepsilon\varepsilon}$. Note that the difference between the asymptotic covariance matrix of the naive estimator and the factor–based risk–premium estimator is $\left(1 - \lambda' \Sigma_{FF}^{-1} \lambda\right) (\Sigma_{\varepsilon\varepsilon} - \beta(\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \beta)^{-1} \beta')$. In order to understand the efficiency gains from adding the information on the factor model, we will next analyse this formula. The following corollary formalizes the relation between the asymptotic covariance matrices of the naive estimator and the factor–model based risk–premium estimator.

Corollary 1.1. *Impose Assumptions 1.1 and 1.2, and consider the moment conditions (1.3.1). Then, we have the following.*

- *If $\lambda' \Sigma_{FF}^{-1} \lambda < 1$, then the limiting variance of the expected (excess) return estimator $\hat{\beta}\hat{\lambda}$ is at most $\Sigma_{R^e R^e}$.*

Corollary 1.1 shows that there may be precision gains for estimating risk premiums from the added information about the factor model if $\lambda'\Sigma_{FF}^{-1}\lambda$ is smaller than one. Note that although $\lambda'\Sigma_{FF}^{-1}\lambda$ can be larger than one mathematically, it is typically smaller than one given the parameters found in empirical research. Observe that in the one–factor case with a traded factor, $\lambda'\Sigma_{FF}^{-1}\lambda$ is the squared Sharpe ratio of that factor. This squared Sharpe ratio is, for stocks and stock portfolios, generally much smaller than 1. Moreover, plugging in the estimates from the Fama–French three factor model (based on GMM with moment conditions (1.3.1)) gives $\lambda'\Sigma_{FF}^{-1}\lambda = 0.06$. Note that the smaller the value for $\lambda'\Sigma_{FF}^{-1}\lambda$, the larger the efficiency gains from imposing a factor model.

As mentioned earlier, we study the empirical relevance of our results by using the parameter values from the FF 3–factor model estimated with FF 25 size–B/M portfolios. In particular, we estimate the parameters by using GMM with the moment conditions (1.3.1). We, then, calculate the (asymptotic) variances of the factor–model based risk–premium estimates for all 25 FF portfolios by plugging the parameter estimates into (1.4.1). Comparing the standard deviation of the factor–model based risk–premium estimators to those of the naive estimators, we see that the factor–model based risk–premium estimators are more precise than the naive estimators. In particular, using the 3–factor model in estimating risk premiums of 25 FF portfolios leads to striking gains in standard deviations with 32% on average over assets.

1.4.2 Precision with Moment Conditions for Traded Factors

When the risk factors are traded, meaning that the factor is an excess return, additional restrictions on the prices of risk can be incorporated into the estimation. With the availability of such information, one could expect efficiency gains in estimating both the prices of risk and the expected (excess) returns. In this section, we consider such case and derive the asymptotic variances of the expected (excess) return estimators with the moment conditions for the case all factors are traded.

Theorem 1.4. *Suppose that all factors are traded. Under Assumptions 1.1 and 1.2, consider the moment conditions (1.3.2)*

$$E[h_t(\alpha,\beta,\lambda)] = E\begin{bmatrix}\begin{bmatrix}1\\F_t\end{bmatrix}\otimes[R_t^e-\alpha-\beta F_t]\\ F_t^e-\lambda\end{bmatrix}=0.$$

Then, the limiting variance of the expected (excess) return estimator $\hat{\beta}\hat{\lambda}$ is given by

$$\Sigma_{R^eR^e} - \left(1-\lambda'\Sigma_{FF}^{-1}\lambda\right)\Sigma_{\varepsilon\varepsilon}. \tag{1.4.2}$$

The theorem above shows that when the factors are traded, the asymptotic covariance matrices of the factor–based risk–premium estimators may change. This is because we incorporate, in the estimation, the restriction that prices of risk associated with factors equal to the expected return of that factor.

Theorem 1.4 allows us to study the efficiency gains for estimating risk premiums from a model where the factors are traded compared to historical averages. Comparing the asymptotic covariance matrix of the factor–based risk–premium estimators from GMM (1.3.2) to the one of the naive estimator, we observe that the difference is given by $\left(1-\lambda'\Sigma_{FF}^{-1}\lambda\right)\Sigma_{\varepsilon\varepsilon}$. Moreover, observe that asymptotic covariance matrix of risk–premium estimator based on GMM with (1.3.2) can be different from the ones of the risk–premium estimator based on GMM with (1.3.1), which indicates that there may be efficiency gains from the information about the factors being traded. The following corollary formalizes these issues.

Corollary 1.2. *Suppose that all factors are traded. Under Assumption 1.1 and 1.2, consider the GMM estimator based on the moment conditons (1.3.2). Then, we have the following.*

1. *If $\lambda'\Sigma_{FF}^{-1}\lambda < 1$, then the limiting variance of the expected (excess) return estimator $\hat{\beta}\hat{\lambda}$ is at most $\Sigma_{R^eR^e}$.*

2. *The limiting variance of this expected (excess) return estimator is at most the limiting variance of the expected (excess) return estimator based on the moment conditions (1.3.1).*

Plugging in the parameter estimates from the analysis of Fama–French model gives $\lambda'\Sigma_{FF}^{-1}\lambda < 1 = 0.05$. Note that $\lambda'\Sigma_{FF}^{-1}\lambda < 1$ is equal to 0.06 in the general case based on GMM 1.3.1. This happens because estimation based on GMM with the set of moment conditions 1.3.1 leads to λ estimates which are different than λ estimates obtained with GMM with 1.3.2. Comparing the standard deviations of the risk–premium estimates based on GMM with (1.3.2) to those of the naive estimators, we see that the risk–premium estimates based on GMM with (1.3.2) typically have smaller asymptotic standard deviations than the naive estimators. In particular, the size of efficiency gains in standard deviations is striking with 36% on average (over assets). Moreover, consistent with Theorem 1.2, the standard deviations of risk–premium estimates based on GMM with (1.3.1) typically exceed those of the naive estimator. Specifically, the risk–premium estimates based on GMM with (1.3.1) have, on average, 16% larger standard deviations than the risk–premium estimates based on GMM with (1.3.2). Overall, there are indeed sizeable precision gains from estimating risk premiums based on factor models based on two sources. First, the linear relation implied by asset pricing models is valuable information in the estimation of risk premiums. Second, when the factors are traded, the additional information that the prices of risk factors equal expected returns of the factors increases the preciseness of risk–premium estimates.

1.4.3 Precision with Moment Conditions Using Factor–Mimicking Portfolios

One may hope that replacing factors by factor–mimicking portfolios may bring efficiency gains since the additional restriction on the price of the factor risk can be incorporated into the estimation. In this subsection, we derive the asymptotic variances of expected (excess) return estimators obtained with factor–mimicking portfolios.

Theorem 1.5. *Under Assumption 1.1 and 1.2, consider the GMM estimator based on the moment conditions (1.3.4)*

$$E\left[h_t(\alpha^m,\beta^m,\Phi_0,\Phi,\lambda^m)\right]=E\begin{bmatrix}\begin{bmatrix}1\\ R_t^e\end{bmatrix}\otimes\left[F_t-\Phi_0-\Phi R_t^e\right]\\ \begin{bmatrix}1\\ F_t^m\end{bmatrix}\otimes\left[R_t^e-\alpha^m-\beta^m F_t^m\right]\\ \Phi R_t^e-\lambda^m\end{bmatrix}=0.$$

Then, the limiting variance of the expected (excess) return estimator, $\hat{\beta}^m\hat{\lambda}^m$, is given by

$$\Sigma_{R^eR^e}-\left(1-\lambda'\left(\beta'\Sigma_{R^eR^e}^{-1}\beta\right)\lambda\right)\left(\Sigma_{R^eR^e}-\beta\left(\beta'\Sigma_{R^eR^e}^{-1}\beta\right)\beta'\right). \tag{1.4.3}$$

Theorem 1.5 enables us to study the efficiency gains in risk premiums using factor–mimicking portfolios. Observe that the difference between the asymptotic covariance matrix of the risk–premium estimator based on GMM with (1.3.4) and the asymptotic covariance matrix of the naive estimator is given by $\left(1-\lambda'\left(\beta'\Sigma_{R^eR^e}^{-1}\beta\right)\lambda\right)$ $\left(\Sigma_{R^eR^e}-\beta\left(\beta'\Sigma_{R^eR^e}^{-1}\beta\right)\beta'\right)$. Note that efficiency gains are dependent on this quantity being positive semi–definite or not. Although we haven't found an answer to this yet, the results from our empirical analysis with FF model illustrates that there is considerable efficiency gains over naive estimation. In particular, estimating risk premiums with GMM (1.3.4) leads to, on average, 32% smaller standard deviations than estimating them with naive estimator. The gains in standard deviation is about 2% compared to the case where risk premiums are estimated with GMM (1.3.1). Note that small gains are expected in our particular empirical example because the three factors of Fama French are zero investment portfolios themselves.

Note that one important difference between Theorem 1.5 and Theorem 1.4 potentially comes from the estimation of the mimicking portfolio weights. The estimation of the weights of the factor–mimicking portfolio potentially leads to different (intuitively higher) asymptotic variances for the betas of the mimicking factors as well as for the mimicking factor prices of risk, and the risk premiums, which are essentially a multiplication of β^m and λ^m. Such issue is similar to errors–in–variables type of corrections in two step Fama–Macbeth estimation, i.e. Shanken (1992) correction in asymptotic variances for generated regressors. We should recall here that GMM standard errors automatically accounts for such effects as it solves the system of moment conditions simultaneously. In particular, in our setting with moments conditions (1.3.4), GMM treats the moments producing Φ simultaneously with the moments generating β^m and λ^m. Hence, the long run covariance matrix captures the effects of estimation of Φ on the standard errors of the β_m and λ^m, hence the risk premiums.

If we consider the Fama–French three factor model with 25 FF–portfolios, we can also intuitively gain insights about the difference between the inferences about risk premiums based on GMM with the two sets of moment conditions (1.3.2) and (1.3.4). In fact, since the factors are traded factors, meaning that they are excess returns themselves, we can estimate the risk premiums via the second set of moment conditions (1.3.2). Moreover, we can also estimate such system via the third set of moment conditions (1.3.4), which has the additional burden of estimating the coefficients for the construction of the mimicking portfolio. Accordingly, GMM estimation via the second set and the third set of moment conditions may lead to different precisions for the risk premium estimates. Table 1.2 presents the efficiency gains in estimating risk premiums of 25 FF portfolios for three sets of moment conditions from an empirical analysis with 25 FF assets employing Fama French 3 factor model. Comparing the second and third columns, we can observe that risk premium estimates employing factor mimicking portfolios, i.e. based on (1.3.4), are less precise than risk premium estimates based on (1.3.2).

1.5 Risk Premium Estimation with Omitted Factors

The asymptotic results in the previous section are based on the assumption that the pricing model is correctly specified. The researcher is assumed to know the true factor model that explains expected excess returns on the assets. In that case, the risk–premium estimators are consistent certainly under our maintained assumption of independently and identically distributed returns. However, the pricing model may be misspecified and this might induce inconsistent risk–premium estimates. We investigate this issue and its solution in the present section.

We consider model misspecification due to ommitted factors. An example of such type of misspecification would be to use Fama–French three factor model if the true pricing model is the four factor Fama–French–Carhart Model. Formally, assume that excess returns are generated by a factor model with two different sets of distinct factors, F and G such that

$$R^e = \alpha^* + \beta^* F + \delta^* G + \varepsilon^* \tag{1.5.1}$$

where ε^* is a vector of residuals with mean zero and $\mathrm{E}\left[F\varepsilon^{*\prime}\right] = 0$ and $\mathrm{E}\left[G\varepsilon^{*\prime}\right] = 0$. Note that the sets of factors F and G perfectly explain the expected excess returns of the test assets, i.e. $\mathrm{E}\left[R^e\right] = \beta^* \lambda_F + \delta^* \lambda_G$.

However, a researcher may be ignorant about the presence of the factors G and thus estimates the model only with the set of factors, F,

$$R^e = \alpha + \beta F + \varepsilon \tag{1.5.2}$$

with ε has mean–zero and $\mathrm{E}\left[F\varepsilon'\right] = 0$ and estimates the exposures, β and the prices of risk λ by incorrectly specifying $\mathrm{E}\left[R^e\right] = \beta\lambda$. Although the researcher might not know the underlying factor model exactly, she allows for misspecification by adding an N-vector of constant terms in estimation, α as in Fama and French (1993).

The asymptotic bias in the parameter estimates for, α, β and λ are presented in the following theorem:

Theorem 1.6. *Assume that returns are generated by (1.5.1) but α, β and λ are estimated from (1.5.2) with GMM (1.3.1). Then,*

1. *$\hat{\alpha}$ converges to $\alpha^* + (\beta^* - \beta)E[F] + \delta^* E[G]$,*

2. *$\hat{\beta}$ converges to $\beta^* + \delta^* Cov\left[G, F^T\right] \Sigma_F^{-1}$,*

3. *$\hat{\lambda}$ converges to $\lambda_F + (\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \beta)^{-1} \beta' \Sigma_{\varepsilon\varepsilon}^{-1} \left[(\beta^* - \beta)\lambda_F + \delta^* \lambda_G\right]$*

in probability.

The Lemma 1.6 shows that, if a researcher ignores some risk factors G, then the risk price estimators associated with factors F are inconsistent if and only if

$$\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \left[(\beta^* - \beta)\lambda_F + \delta^* \lambda_G\right] \neq 0.$$

It is important to note that the inconsistency of the estimates of risk prices may be caused not only by the risk prices of omitted factors but also the bias in betas of the factors F. This result has an important implication: even if the ignored factors are associated with risk prices of zero, the cross–sectional estimates of the prices of risk on the true factors included in the estimation (F) can still be asymptotically biased. This happens in case F and G are correlated, which is often the case.

Next, we analyse the asymptotic bias in the parameter estimates for again, α, β and λ but this time, in case the factors are traded and the estimation is based on GMM with moment conditions (1.3.2) of Section 1.3.2:

Theorem 1.7. *Assume that returns are generated by (1.5.1) but α, β and λ are estimated from (1.5.2) with GMM (1.3.2). Then,*

1. $\hat{\alpha}$ *converges to* $\alpha^* + (\beta^* - \beta)\lambda_F + \delta^*\lambda_G$,

2. $\hat{\beta}$ *converges to* $\beta^* + \delta^* Cov\left[G, F^T\right]\Sigma_F^{-1}$,

3. $\hat{\lambda}$ *converges to* λ_F,

in probability.

Theorem 1.7 illustrates that, even if the researcher forgets some risk factors, risk price estimators will still be asymptotically unbiased. Notice that this is in contrast with the estimator based on GMM with moment conditions (1.3.1) of Section 1.3.1. It is important to note that, if the forgotten factors, G, are uncorrelated with the factors, then the bias in β disappears. Moreover, if the ignored factors are associated with zero prices of risk and uncorrelated with F, then the $\hat{\alpha}$ will converge to zero.

What happens to the risk–premium estimators on individual assets or portfolios if some true factors are ignored? The following corollary provides the consistency condition for risk–premium estimators of individual assets or portfolios.

Corollary 1.3. *If the returns are generated by (1.5.1) and*

- *the model (1.5.2) is estimated with GMM (1.3.1), then the vector of resulting risk–premium estimators* $\hat{\beta}\hat{\lambda}$ *converges to* $E[R^e]$ *if and only if* $[I_N - \beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta'\Sigma_{\varepsilon\varepsilon}^{-1}]E[R^e] = 0$.

- *all factors are traded. If the model (1.5.2) is estimated with GMM (1.3.2), then the vector of resulting risk–premium estimators* $\hat{\beta}\hat{\lambda}$ *converges to* $E[R^e]$ *if and only if* $(\beta^* - \beta)\lambda_F + \delta^*\lambda_G = 0$.

In the view of the theorem above, if the model (1.5.2) is estimated with GMM (1.3.1), the consistency of the risk–premium estimators is dependent on a specific condition that may not be satisfied. Moreover, if the factors are traded and the estimation is via GMM with moment conditions (1.3.2), then the risk–premium estimator obtained may be biased.

In order to capture misspecification, it is a common approach to add an N–vector of constant terms, α, to the model as in (1.5.2). In the following thoerem, we will show that in case of traded factors, it is possible to achieve the consistency for estimating risk premiums.

Theorem 1.8. *Assume that all factors in F are traded. If the returns are generated by (1.5.1) but the model (1.5.2) is estimated with GMM (1.3.2) where the risk price estimates are given by the factor averages, then the estimator $\hat{\alpha} + \hat{\beta}\hat{\lambda}$ is consistent for $E[R^e]$. However, the asymptotic variance of such estimator equals $\Sigma_{R^eR^e}$.*

Theorem 1.8 shows that when all the factors in the estimation (F) are traded and if the estimation is based on GMM with moment conditions (1.3.2), then we obtain a consistent estimator for risk premiums by adding an estimator for the N–vector of constant terms, $\hat{\alpha}$, to $\hat{\beta}\hat{\lambda}$. However, this estimator is not asymptotically more efficient than the naive estimator of risk premiums.

Some asset pricing studies add a one dimensional constant, henceforth λ_0, to the asset pricing specification of expected returns as in $\mathrm{E}[R^e] = 1_N\lambda_0 + \beta\lambda$, where 1_N is an N–vector of ones and make inferences about it. At this stage, we do not analyze the role of such objects. Recall that here α is an N–vector of constants; it does not represent a one dimensional object as λ_0.

It is important to note that adding the $\hat{\alpha}$ to $\hat{\beta}\hat{\lambda}$ does not solve the inconsistency problem if the system is estimated via GMM with (1.3.1). If some factors are non–traded and the parameters are estimated via GMM with (1.3.1), adding the $\hat{\alpha}$ capturing the misspecification to $\hat{\beta}\hat{\lambda}$ doesn't lead to consistent estimates of $\mathrm{E}[R^e]$. In particular, $\hat{\alpha}+\hat{\beta}\hat{\lambda}$ converges to $\mathrm{E}[R^e]-\beta(\lambda-\mathrm{E}[F])$ and $\lambda-\mathrm{E}[F]$ is not necessarily zero.

1.6 Application: Portfolio Choice with Parameter Uncertainty

This section analyzes the performances of portfolios based on different risk–premium estimates in the optimization problem of Markowitz (1952). The implementation of the mean–variance framework of Markowitz (1952) requires the estimation of first two moments of the asset returns. Mean–variance portfolios could be constructed by plugging in both factor–based risk–premium estimates or historical averages. Because we showed in previous sections that factor–model based risk–premium estimators are more precise than the naive estimator, the following questions arise: how is the performance of the mean–variance portfolio affected by the improvement in the precision of risk–premium estimates? To answer this, we analyze, in this section, the out of sample performances of the tangency portfolios based on the various risk–premium estimators in a simulation analysis.

Optimization Problem: Suppose a risk–free asset exists and w is the vector of relative portfolio allocations of wealth to N risky assets. The investor has preferences that are fully characterized by the expected return and variance of his selected portfolio, w. The investor maximizes his expected utility, by choosing the vector of portfolio weights w,

$$\mathrm{E}\left[U\right] = w'\mu^e - \frac{\gamma}{2} w'\Sigma_{RR} w, \tag{1.6.3}$$

where γ measures the investor's risk aversion level, μ^e and Σ_{RR}[3] denote the expected excess returns on the assets and covariance matrix of returns. The solution to the maximization problem above is given by $w_{opt} = \frac{1}{\gamma}\Sigma_{RR}\mu^e$. From this expression, the vector of tangency portfolio weights can be derived by incorporating the constraint

[3]Note that Σ_{RR}=$\Sigma_{R^eR^e}$.

that portfolio weights of risky assets sum to one and is given by[4]

$$w_{tg} = \frac{\Sigma_{RR}\mu^e}{\iota_N' \Sigma_{RR}\mu^e}, \tag{1.6.4}$$

where ι_N is an N–vector of ones.

In the optimization problem above, since the true risk premium vector, μ^e, and the true covariance matrix of asset returns, Σ_{RR}, are unknown, in empirical work, one needs to estimate them. Following the classical "plug in" approach, the moments of the excess return distribution, μ^e and Σ_{RR}, are replaced by their estimates.

Portfolios Considered: We consider four portfolios constructed with different risk–premium estimators: the tangency portfolio constructed with historical averages, the tangency portfolio constructed with the factor–model based estimates (GMM–Gen, GMM–Tr, GMM-Mim). Note that the covariance matrix is estimated using the traditional sample counterpart, $1/(T-1)\sum_1^T (R_t - \bar{R}_t)(R_t - \bar{R}_t)'$, where $\bar{R}_t$ is the sample average of returns. We also consider the global minimum variance portfolio[5] to which we compare the performance of the portfolios based on the risk–premium estimates. Note that the implementation of this portfolio only requires estimation of the covariance matrix, for which we again use the sample counterpart, and completely ignores the estimation of expected returns.

Performance Evaluation Criterion and Methodology: We compare performances of the portfolios considered by using out-of-sample Sharpe Ratios. We set an initial window length over which we estimate the mean vector of excess returns and covariance matrix, and obtain the various portfolio weights. For our analysis, the initial window length is of 120 data points, corresponding to 10 years of data. We then calculate the one-month ahead returns, $\hat{w}_t R^e_{t+1}$, of the estimated portfolios. Next, we reestimate the portfolio weights by including the next month's return

[4]Because it lies on the mean variance frontier.

[5]This portfolio is obtained by minimizing the portfolio variance with respect to the weights with the only constraint that weights sum to 1 and the N–vector of portfolio weights is given by $w_{gmv} = \Sigma_{RR}\iota_N / \iota_N \Sigma_{RR}\iota_N$

and use this to calculate the return for the subsequent month. We continue doing this and obtain the time series of out–of–sample excess returns for each portfolio considered, from which we calculate the out–of–sample Sharpe ratios.

Simulation Experiment: We consider twenty–five Fama and French (1992) portfolios sorted by size and book–to–market as risky assets and the nominal 1–month Treasury bill rate as a proxy for risk–free rate (both available on French's website). We use the 3 Fama and French (1992) portfolios (market, book–to–market and size factors) as our factors. To make our simulations realistic, we calibrate the parameters by using the monthly data of the aforementioned portfolios, from January 1963 until December 2012. Specifically, we estimate $\alpha, \beta, \mu_F, \Sigma_{FF}, \Sigma_{\varepsilon\varepsilon}, \lambda$ and take them to be the truth in the simulation exercise to generate samples of 597 observations. To be precise, we use the following return–generating process:

$$R_t^e = \alpha + \beta F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1.6.5}$$

with F_t and ε_t drawn from multivariate normal distributions with the true moments. Note that we set α equal to zero for all simulations. We simulate independent sets $Z = 5,000$ return samples with the full sample size of 597. For each set of simulated sample, we calculate the out–of–sample Sharpe ratios for the various portfolios.

Table 1.3 provides the simulation results for the out–of–sample Sharpe ratios of different portfolios. In particular, we provide results on the tangency portfolios based on different risk–premium estimates and global minimum variance portfolios. Moreover, we provide the true Sharpe ratio of the tangency portfolio, which we refer as theoretical. For each portfolio, we present the average estimate over simulations, $\overline{SR}$, the bias as the percentage of the population Sharpe ratios, $(\overline{SR} - SR)/SR$ and the root–mean–square error(RMSE) in parantheses, the square root of $\sum_{s=1}^{Z}(\hat{SR}_s - SR)/Z$, where $Z = 5,000$.

In order to isolate the effect of the error in risk–premium estimates, we present our results with true and estimated Σ_{RR}. Firstly, note that the true Sharpe ratio of the tangency portfolio is superior to the portfolios based on estimated risk–premiums or covariance matrix of asset returns. Comparing the average Sharpe ratio of the tangency portfolio based on historical averages to the true Sharpe ratio of tangency portfolio, we see that the bias is striking and negative with −56.26% and −56.88%, depending on the covariance matrix of asset returns is the true one or the estimated one. However, using factor–models to estimate risk–premiums reduces the bias in Sharpe ratios substantially to a level ranging from −18.02% to −26.25%. In particular, with GMM–Gen estimates, average Sharpe ratio of the tangency portfolio is 0.1541 in case of true covariance matrix (with an improvement of 69% over the average Sharpe ratios with the historical averages) and 0.1670 in case of an estimated covariance matrix (with an improvement of 90% over the average Sharpe ratios with the historical averages). Among the tangency portfolios constructed with factor–model based risk–premium estimates, GMM–Tr estimates perform, in terms of bias, the best given that the covariance matrix is known and GMM–Gen estimates perform, in terms of bias, the best given that the covariance matrix is estimated. However, the differences in biases are minimal for all tangency portfolios constructed with factor–model based risk–premium estimators.

Next, we analyse the RMSEs of the various portfolios. Out–of–sample Sharpe ratio of the tangency portfolios based on historical averages is extremely volatile across simulations. That is, it has a RMSE of 0.1353 (given the average estimate 0.0879) if the covariance matrix is estimated. However, using factor–based risk–premium estimators decreases the RMSEs substantially. Among the tangency portfolios based on factor–model based risk–premium estimators, GMM–Tr performs the best with a RMSE of 0.0881 (given the average estimate of 0.1668), as expected from our asymptotic analyses of risk–premium estimators in previous sections. However, the differences in RMSEs are minor among the portfolios with factor–based risk–premium estimates.

We also compare the performance of the tangency portfolio estimates to the global minimum variance portfolios. Jagannathan and Ma (2003) and DeMiguel et al. (2009) note that the estimation error in expected returns is so large that focusing on the minimum variance portfolios, which ignore the expected returns completely, is less sensitive to the estimation error than the mean–variance portfolios. In particular, it has been shown in empirical studies that minimum–variance portfolios usually has better out–of–sample performance than any other mean–variance portfolios.[6] Consistent with them, we find that global minimum variance portfolio with an estimated covariance matrix has a higher average Sharpe ratio, 0.0984 and substantially lower RMSE, 0.0469, compared to the tangency portfolios constructed with historical averages. However, the average Sharpe ratios of the tangency portfolios are considerably larger than the average Sharpe ratio of the global minimum variance portfolio when the factor–based risk–premium estimates are used. Specifially, average Sharpe ratios of the GMM–Gen, GMM–Tr, GMM–Mim are 0.1670, 0.1668, 0.1670 respectively. Overall, using the factor–model based risk–premium estimators improves the performance of tangency portfolios substantially over the plug in estimates of historical averages, in terms of both bias and RMSEs. Moreover, in contrast to the tangency portfolios with historical averages, these portfolios perform considerably better than the global minimum variance portfolio.

1.7 Conclusions

It has been the standard technique in the literature to use average historical returns as estimates of expected excess returns, that is risk premiums, on individual assets or portfolios. However, the finance literature provides a wide variety of risk–return models which imply a linear relationship between the expected excess returns and their exposures.

In this paper, we show that, when correctly specified, such parametric specifica-

[6] Compare DeMiguel et al. (2009), Jagannathan and Ma (2003), and Jorion (1991).

tions on the functional form of risk premiums lead to significant inference gains for estimating expected (excess) returns. Moreover, we show that using a misspecified asset pricing model in the sense that some factors are forgotten generally leads to inconsistent estimates. However, in case the factors are traded, then adding an alpha to the model capturing mispricing leads to consistent estimators.

Out of sample performance of tangency portfolios significantly improves if factor–based estimates of risk premium are used in portfolio weights instead of the classical historical averages.

1.8 Tables

Table 1.1: Efficiency gains for factor–based risk premium estimators

This table presents the average gains in standard deviations for the various risk premium estimates. The test assets are the 25 Fama–French size and book–to–market portfolios and the factors are the three factors of Fama French (1992). The first row illustrates the gains for three different factor–model based risk–premium estimates (GMM–Gen, GMM–Tr and GMM–Mim) over the historical averages. The table presents the average gains over 25 assets.

	RP with GMM–Gen	RP with GMM–Tr	RP with GMM–Mim
Naive	0.32	0.36	0.32
RP with GMM-Gen	-	0.16	0.02

Table 1.2: Efficiency gains for factor–model based risk premium estimates for 25 FF assets

This table illustrates the gains in standard deviations for the various risk premium estimates for 25 portfolios formed by Fama–French (1992,1993). The factors are the three factors (market, size and book–to–market) of Fama–French (1992). The results are based on monthly data from January 1963 until October 2012, i.e. 597 observations for each portfolio. The first column (Gen–N) presents the gains of the factor–model based estimates of risk premiums based on GMM with 1.3.1 over the naive estimate of historical averages. The second and third columns present the gains of factor model based estimates of risk premiums based on GMM with 1.3.2 and with 1.3.4 over naive estimates respectively. Fourth column corresponds to the gains from estimating the system based on GMM with 1.3.2 over the case of estimating the system based on GMM with 1.3.2. The last column presents the gains from making use of mimicking portfolios and estimate the system with GMM (1.3.4) over estimation with GMM (1.3.1).

Assets	Gen–N	Tr–N	Mim–N	Tr–Gen	Mim–Gen
1	0.4374	0.5305	0.4390	0.2750	0.0339
2	0.4269	0.4980	0.4284	0.2359	0.0331
3	0.3404	0.4001	0.3416	0.1990	0.0264
4	0.3179	0.3665	0.3190	0.1738	0.0246
5	0.2348	0.2933	0.2356	0.1711	0.0182
6	0.2751	0.3856	0.2759	0.2606	0.0213
7	0.2536	0.3313	0.2544	0.2067	0.0196
8	0.2169	0.2797	0.2176	0.1725	0.0168
9	0.2394	0.2714	0.2402	0.1243	0.0185
10	0.2250	0.2534	0.2257	0.1138	0.0174
11	0.2808	0.3668	0.2817	0.2271	0.0218
12	0.2705	0.3125	0.2714	0.1510	0.0209
13	0.3044	0.3239	0.3054	0.1059	0.0236
14	0.3005	0.3174	0.3015	0.0978	0.0233
15	0.3127	0.3299	0.3137	0.1006	0.0242
16	0.3174	0.3542	0.3184	0.1498	0.0246
17	0.3234	0.3385	0.3244	0.0951	0.0250
18	0.3458	0.3619	0.3470	0.1009	0.0268
19	0.3244	0.3507	0.3254	0.1269	0.0251
20	0.3713	0.3935	0.3725	0.1221	0.0288
21	0.2257	0.2460	0.2264	0.0953	0.0175
22	0.3368	0.3615	0.3379	0.1245	0.0261
23	0.4259	0.4640	0.4274	0.1694	0.0330
24	0.3868	0.4591	0.3881	0.2308	0.0300
25	0.5039	0.5596	0.5058	0.2173	0.0390

Table 1.3: Tangency out–of–sample Sharpe ratio estimates with different risk–premium estimates

This table provides, average sharpe ratio estimate over simulations, its percentage error, compared to the true sharpe ratio and the RMSE(in paranthesis) of the sharpe ratios constructed with various mean estimates. Note that the variance covariance matrix is estimated by the sample variance covariance matrix.

	True Σ_{RR}	Estimated Σ_{RR}	Theoretical
True μ^e	0.2129	0.2038	0.2090
	0.0188	-0.0249	
Naive	0.0914	0.0879	0.2090
	-0.5626	-0.5688	
	(0.1375)	(0.1353)	
GMM–Gen	0.1541	0.1670	0.2090
	-0.2625	-0.1802	
	(0.0977)	(0.0867)	
GMM–Tr	0.1552	0.1668	0.2090
	-0.2574	-0.1814	
	(0.0966)	(0.0881)	
GMM–Mim	0.1541	0.1670	0.2090
	-0.2625	-0.1802	
	(0.0966)	(0.0867)	
GMV	0.0974	0.0984	0.2090
	(0.0461)	(0.0469)	

Appendix

1.A Proofs for Chapter 1

In the rest of the paper, the covariance matrix of the factor–mimicking portfolios is denoted by $\Sigma_{F^m F^m}$.

1.A.1 Equivalence of factor pricing using mimicking portfolios

Proof of Theorem 1.1. Define M^m as the projection of M onto the augmented span of excess returns,

$$M^m = \mathbb{P}(M|1, R^e) \tag{1.A.1}$$

so that

$$\begin{aligned} \mathrm{E}\left[M\right] &= \mathrm{E}\left[M^m\right], & (1.A.2) \\ \mathrm{Cov}\left[M, R^e\right] &= \mathrm{Cov}\left[M^m, R^e\right]. & (1.A.3) \end{aligned}$$

Thus, we have

$$\begin{aligned}
\beta\lambda &= \mathrm{Cov}\left[R^e, F'\right]\Sigma_{FF}^{-1}\left(-\frac{1}{\mathrm{E}\left[M\right]}\Sigma_{FF}b\right) \\
&= -\frac{1}{\mathrm{E}\left[M\right]}\mathrm{Cov}\left[R^e, F'\right]b \\
&= -\frac{1}{\mathrm{E}\left[M^m\right]}\mathrm{Cov}\left[R^e, F^{m\prime}\right]b \\
&= -\frac{1}{\mathrm{E}\left[M^m\right]}\mathrm{Cov}\left[R^e, F^{m\prime}\right]\Sigma_{F^mF^m}^{-1}\Sigma_{F^mF^m}b \\
&= \beta^m\lambda^m,
\end{aligned} \tag{1.A.4}$$

which completes the proof. ■

1.A.2 Precision of Parameter Estimators Given a Factor Model

This section provides the proofs for asymptotic properties of the parameter estimators under the specified linear factor model. The lemma 1.9 below illustrates the asymptotic distribution of the GMM estimators with a given set of moment conditions provided that a pre–specified matrix A, that essentially determines the weigths of the overidentifying moments, is introduced. Thereafter, these results will be used to calculate the variance covariance matrix for the moment conditions (1.3.1), (1.3.2) and (1.3.4), respectively. Under appropriate regularity conditions, see, e.g., Hall (2005), Chapter 3.4, we have the following result:

Lemma 1.9. *Let $\theta \in \mathbb{R}^p$ be a vector of parameters and the moment conditions are given by $E[h_t(\theta)] = 0$ where $h_t(\theta) \in \mathbb{R}^q$, independently and identically distributed over time. Given a prespecified matrix $A \in \mathbb{R}^{p\times q}$, its consistent estimator $\hat{A}$ and $\hat{A}\frac{1}{T}\sum_{t=1}^{T} h_t(\hat{\theta}) = 0$,*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, [AJ]^{-1}ASA'[J'A']^{-1}\right), \tag{1.A.5}$$

where,

$$J = E\left[\frac{\partial h_t(\theta)}{\partial \theta'}\right], \tag{1.A.6}$$

$$S = E[h_t(\theta)h_t(\theta)']. \tag{1.A.7}$$

The above lemma presents the asymptotic distribution of the parameters in a general GMM context. In the subsequent lemmas, limiting distributions for the expected (excess) return estimators based on the moment conditions (1.3.1), (1.3.2) and (1.3.4), respectively.

Lemma 1.10. *Under Assumptions 1.1, 1.2 and the moment conditions (1.3.1) with parameter vector* $\theta = (\alpha', vec(\beta)', \lambda')'$, *we have*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V), \tag{1.A.8}$$

with

$$V = \begin{bmatrix} \begin{bmatrix} 1 + \mu_F' \Sigma_{FF}^{-1} \mu_F & -\mu_F' \Sigma_{FF}^{-1} \\ -\Sigma_{FF}^{-1} \mu_F & \Sigma_{FF}^{-1} \end{bmatrix} \otimes \Sigma_{\varepsilon\varepsilon} & V_c \\ V_c' & (1 + \lambda' \Sigma_{FF}^{-1} \lambda)(\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \beta)^{-1} + \Sigma_{FF} \end{bmatrix}$$

where $\mu_F = E[F_t]$ *and* $V_c = \begin{bmatrix} 1 + \mu_F' \Sigma_{FF}^{-1} \lambda \\ -\Sigma_{FF}^{-1} \lambda \end{bmatrix} \otimes \beta(\beta' \Sigma_{\varepsilon\varepsilon}^{-1} \beta)^{-1}$.

Proof. The proof follows from plugging the appropriate matrices for the moment conditions provided in Section 1.3.1 into the variance covariance formula in (1.A.5) and performing the matrix multiplications. Below, we provide the limiting variance covariance matrix (S) and the Jacobian (J) for this specific set of moment conditions,

$$S = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \mu_F' \otimes \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\varepsilon} \\ \mu_F \otimes \Sigma_{\varepsilon\varepsilon} & [\Sigma_{FF} + \mu_F \mu_F'] \otimes \Sigma_{\varepsilon\varepsilon} & \mu_F \otimes \Sigma_{\varepsilon\varepsilon} \\ \Sigma_{\varepsilon\varepsilon} & \mu_F' \otimes \Sigma_{\varepsilon\varepsilon} & \beta \Sigma_{FF} \beta' + \Sigma_{\varepsilon\varepsilon} \end{bmatrix},$$

$$J(\theta) = \mathrm{E}\left[\frac{\partial h_t(\theta)}{\partial \theta'}\right] = \begin{bmatrix} -\begin{bmatrix} 1 & \mu_F' \\ \mu_F & \Sigma_{FF} + \mu_F \mu_F' \end{bmatrix} \otimes I_N & 0_{N(K+1)\times K} \\ \begin{bmatrix} 0_{N\times N} & -\lambda' \otimes I_N \end{bmatrix} & -\beta \end{bmatrix}.$$

Furthermore

$$A = \begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)\times N} \\ 0_{K\times N(K+1)} & \beta' \Sigma_{\varepsilon\varepsilon}^{-1} \end{bmatrix},$$

so that the limiting variance of GMM estimator for θ is obtained by performing the matrix multiplications $[AJ]^{-1} ASA' [J'A']^{-1}$. ■

Lemma 1.11. *Suppose that all factors are traded. Then, under Assumptions 1.1, 1.2 and the moment conditions (1.3.2) with parameter vector $\theta = (\alpha', vec(\beta)', \lambda')'$, we have*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V), \tag{1.A.9}$$

with

$$V = \begin{bmatrix} \begin{bmatrix} 1 + \mu_F' \Sigma_{FF}^{-1} \mu_F & -\mu_F' \Sigma_{FF}^{-1} \\ -\Sigma_{FF}^{-1} \mu_F & \Sigma_{FF}^{-1} \end{bmatrix} \otimes \Sigma_{\varepsilon\varepsilon} & 0_{N(K+1)\times K} \\ 0_{K\times N(K+1)} & \Sigma_{FF} \end{bmatrix}.$$

Proof. The proof follows from plugging the appropriate matrices for the moment conditions (1.3.2) into the variance covariance formula in (1.A.5) and performing the matrix multiplications. Below, we provide the limiting variance covariance matrix

(S), Jacobian (J) for this specific set of moment conditions, In this case,

$$S = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \mu_F' \otimes \Sigma_{\varepsilon\varepsilon} & 0_{N\times K} \\ \mu_F \otimes \Sigma_{\varepsilon\varepsilon} & [\Sigma_{FF} + \mu_F\mu_F'] \otimes \Sigma_{\varepsilon\varepsilon} & 0_{NK\times K} \\ 0_{K\times N} & 0_{K\times NK} & \Sigma_{FF} \end{bmatrix},$$

and

$$J(\theta) = \begin{bmatrix} -\begin{bmatrix} 1 & \mu_F' \\ \mu_F & \Sigma_{FF} + \mu_F\mu_F' \end{bmatrix} \otimes I_N & 0_{N(K+1)\times K} \\ 0_{K\times N(K+1)} & I_K \end{bmatrix}.$$

Thus, the limiting variance of the GMM estimator for θ is obtained by performing the matrix multiplications $J^{-1}S[J']^{-1}$ since $A = I_{N(K+1)+K}$. ■

The next lemma provides the asymptotic properties of the GMM estimatior with factor–mimicking portfolios.

Lemma 1.12. *Given that Assumption 1.1, 1.2 are satisfied and that (1.2.8)–(1.2.10) hold, then under the moment conditions (1.3.4), for* $\theta = (vec\,(\beta^m)', \lambda^{m\prime})'$, *we have*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V), \tag{1.A.10}$$

with

$$V = \begin{bmatrix} \Sigma_{F^mF^m}^{-1} \otimes \beta^m\beta^{m\prime} & -\Sigma_{F^mF^m}^{-1}\mu_{F^m} \otimes \beta^m \\ -\mu_{F^m}'\Sigma_{F^mF^m}^{-1} \otimes \Sigma_{uu}\beta^{m\prime} & \mu_{R^e}'\Sigma_{R^eR^e}^{-1}\mu_{R^e}\Sigma_{uu} + \Sigma_{F^mF^m} \end{bmatrix}.$$

Proof. The proof follows again from plugging the appropriate matrices for the moment conditions (1.3.4) into the variance covariance formula in (1.A.5) and performing the matrix multiplications. Now, observe that from (1.A.7), we have

$$S=\begin{bmatrix} \begin{bmatrix} 1 & \mu_{R^e}' \\ \mu_{R^e} & \Sigma_{RR}+\mu_{R^e}\mu_{R^e}' \end{bmatrix}\otimes\Sigma_{uu} & 0_{K(1+N)\times N(K+1)} & 0_{K(1+N)\times K} \\ 0_{N(K+1)\times K(1+N)} & \begin{bmatrix} 1 & \mu_{F^m}' \\ \mu_{F^m} & \Sigma_{F^mF^m}+\mu_{F^m}\mu_{F^m}{}' \end{bmatrix}\otimes\Sigma_{\varepsilon^m\varepsilon^m} & 0_{N(K+1)\times K} \\ 0_{K\times K(1+N)} & 0_{K\times N(K+1)} & \Sigma_{F^mF^m} \end{bmatrix},$$

and from (1.A.6), we have

$$J(\theta)=\mathrm{E}\begin{bmatrix} -\begin{bmatrix} 1 & R_t^{e\prime} \\ R_t^e & R_t^eR_t^{e\prime} \end{bmatrix}\otimes I_K & 0_{K(1+N)\times N(K+1)} & 0_{K(1+N)\times K} \\ -\begin{bmatrix} 0 & R_t^{e\prime} \\ 0_{K\times 1} & \Phi(R_t^eR_t^{e\prime}) \end{bmatrix}\otimes\beta^m & -\begin{bmatrix} 1 & F_t^{m\prime} \\ F_t^m & F_t^mF_t^{m\prime} \end{bmatrix}\otimes I_N & 0_{N(K+1)\times K} \\ 0_K \quad R_t^{e\prime}\otimes I_K & 0_{K\times N(K+1)} & -I_K \end{bmatrix},$$

with $A = I_{K(1+N)+N(K+1)+K}$. Thus, the limiting variance of the GMM estimator for $\theta = (\mathrm{vec}\,(\beta^m)', \lambda^{m\prime})'$ is obtained by performing the matrix multiplications $J^{-1}S[J']^{-1}$. Here, it is worth stressing that the limiting variance covariance matrix obtained by performing the matrix multiplications corresponds to the parameter vector

$$(\Phi_0{}', \mathrm{vec}\,(\Phi)', \alpha^{m\prime}, \mathrm{vec}\,(\beta^m)', \lambda^{m\prime})' \tag{1.A.11}$$

Therefore, the asymptotic variance covariance matrix for $\theta = (\mathrm{vec}\,(\beta^m)', \lambda^{m\prime})'$ is the lower-right $KN+K$ by $KN+K$ sub-matrix of the larger variance covariance matrix. ■

Lemmas 1.10–1.12 allow us to study the asymptotic properties of the obtained risk premium estimators. It is worth mentioning that the lower–left NK+K dimensional square matrices of the variance covariance matrices in Lemma 1.10 and 1.11 give the variance covariance matrices corresponding to parameters $(\mathrm{vec}\,(\beta)', \lambda')'$. We will use these results to derive the variance covariance matrices of risk premium estimators in the following section.

Proof of Theorem 1.2. This follows from a direct application of the Central Limit Theorem. ■

Proofs of Theorems 1.3 and 1.4. We are interested in the asymptotic distribution of $g(\beta, \lambda) = \beta\lambda$. Given

$$(\mathrm{vec}\left(\hat{\beta}\right)', \hat{\lambda}')' - (\mathrm{vec}\,(\beta)', \lambda')' \xrightarrow{d} \mathcal{N}(0, V_{\beta,\lambda}), \tag{1.A.12}$$

we have, by applying the delta method, that

$$\sqrt{T}\left(g(\hat{\beta}, \hat{\lambda}) - g(\beta, \lambda)\right) \xrightarrow{d} \mathcal{N}(0, \dot{g}' V_{\beta,\lambda} \dot{g}), \tag{1.A.13}$$

with

$$\dot{g} = \begin{bmatrix} \lambda' \otimes I_N & \beta \end{bmatrix}.$$

Remember that Lemma 1.10 and 1.11 give the asymptotic distributions of $\sqrt{T}(\hat{\theta} - \theta)$ where $\theta = (\alpha', \mathrm{vec}\,(\beta)', \lambda')'$ for the moment conditions (1.3.1) and (1.3.2). Observe that $V_{\beta,\lambda}$ is the lower $NK + K$ block diagonal matrix of the variance covariance matrices provided in Lemma 1.10 and 1.11. Hence, the asymptotic variances of the risk premium estimators in Theorems 1.3 and 1.4 follow from plugging in the limiting variance covariance matrices of $(\mathrm{vec}\,(\beta)', \lambda')'$ and calculating $\dot{g}' V_{\beta,\lambda} \dot{g}$. ■

The following theorem follows from the formula of partitioned inverses.

Lemma 1.13. *Let*

$$\mathbf{K} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

be a symmetric matrix and assume that K_{22}^{-1} exists. Then $K \geq 0$ is equivalent to $K_{22} \geq 0$ and $K_{11} - K_{12}K_{22}^{-1}K_{21} \geq 0$.

Proof of Theorem 1.5. We are interested in $g(\beta^m, \lambda^m) = \beta^m \lambda^m$. Given

$$(\mathrm{vec}\left(\hat{\beta^m}\right)', \hat{\lambda^m}')' - (\mathrm{vec}\left(\beta^m\right)', \lambda^{m\prime})' \xrightarrow{d} \mathcal{N}(0, V_{\beta^m,\lambda^m}), \tag{1.A.14}$$

Then, by applying the delta method, we have

$$\sqrt{T}(g(\hat{\beta^m}, \hat{\lambda^m}) - g(\beta^m, \lambda^m)) \xrightarrow{d} \mathcal{N}(0, \dot{g}' V_{\beta^m,\lambda^m} \dot{g}) \tag{1.A.15}$$

and note that here

$$\dot{g} = \left[\begin{array}{cc} \lambda^{m\prime} \otimes I_N & \beta^m \end{array}\right].$$

Then, we have

$$\dot{g}' V_{\beta^m,\lambda^m} \dot{g} = \lambda^{m\prime} {\Sigma_{F^m F^m}}^{-1} \lambda^m \Sigma_{\varepsilon^m \varepsilon^m} + \beta^m \Sigma_{F^m F^m} \beta^{m\prime} \tag{1.A.16}$$

$$+ \left(\mu_{F^m}' \Phi' \Sigma_{RR}^{-1} \Phi \mu_{F^m} - \lambda^{m\prime} {\Sigma_{F^m F^m}}^{-1} \lambda^m\right) \beta^m \Sigma_{uu} \beta^{m\prime}. \tag{1.A.17}$$

The result follows from plugging the β^m and Φ respectively into the above equation.

■

Proofs of Corollaries 1.1 and 1.2. We need to study the difference between:

1. The limiting variance of the historical averages and the limiting variance of the expected (excess) return estimator based on (1.3.1), referring to Corollary 1.1
2. The limiting variance of the historical averages and the limiting variance of the expected (excess) return estimator based on (1.3.2), referring to the Corollary 1.2
3. The limiting variance of the expected (excess) return estimator based on (1.3.1) and The limiting variance of the expected (excess) return estimator based

on (1.3.2), referring to the Corollary 1.2

Suppose $1 - \lambda'\Sigma_{FF}^{-1}\lambda < 1$ and $1 - \lambda'\beta'\Sigma_{R^eR^e}^{-1}\beta\lambda < 1$. In the following, we will show that the differences between the limiting variance above are positive semi–definite. Lemma 1.13 will be used to establish the positive semi–definiteness of the differences. To prove 1, we need to study the difference

$$\begin{aligned}&\Sigma_{R^eR^e} - \left(\Sigma_{R^eR^e} - \left(1 - \lambda'\Sigma_{FF}^{-1}\lambda\right)\left[\Sigma_{\varepsilon\varepsilon} - \beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta'\right]\right) \\ &= \left(1 - \lambda'\Sigma_{FF}^{-1}\lambda\right)\left[\Sigma_{\varepsilon\varepsilon} - \beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta'\right]\end{aligned} \tag{1.A.18}$$

In order to show that $\Sigma_{\varepsilon\varepsilon} - \beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta'$ is positive semi–definite, we will use Lemma 1.13. Now, define $K_1 = \Sigma_{\varepsilon\varepsilon}^{1/2}$ and $K_2 = \beta'\Sigma_{\varepsilon\varepsilon}^{-1/2}$. Then,

$$\mathbf{K} = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \begin{bmatrix} K_1' & K_2' \end{bmatrix} = \begin{bmatrix} K_1K_1' & K_1K_2' \\ K_2K_1' & K_2K_2' \end{bmatrix}$$

so that

$$\mathbf{K} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \beta \\ \beta' & \beta'\Sigma\varepsilon\varepsilon^{-1}\beta \end{bmatrix}.$$

Then, Lemma 1.13 yields that

$$\Sigma_{\varepsilon\varepsilon} - \beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta' \geq 0 \tag{1.A.19}$$

To prove 2 referring to Corollary 1.2 , we need to study the difference

$$\begin{aligned}&\Sigma_{R^eR^e} - \left(\Sigma_{R^eR^e} - \left(1 - \lambda'\Sigma_{FF}^{-1}\lambda\right)\Sigma_{\varepsilon\varepsilon}\right) \\ &= \left(1 - \lambda'\Sigma_{FF}^{-1}\lambda\right)\Sigma_{\varepsilon\varepsilon}\end{aligned} \tag{1.A.20}$$

Since $\Sigma_{\varepsilon\varepsilon}$ is positive semi-definite, the first part of the Corollary 1.2 follows. In order to prove the second part of the Corollary 1.2, we need to show that $\beta(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta)^{-1}\beta'$

is positive semi-definite. Since $\Sigma_{\varepsilon\varepsilon}$ is positive semi–definite, $\Sigma_{\varepsilon\varepsilon}^{-1}$ is also positive semi–definite. There exists a positive semi–definite matrix Z such that $Z^2 = \Sigma_{\varepsilon\varepsilon}^{-1}$. Then, $((Z\beta)'(Z\beta)) = \beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta$ and it is positive semi–definite. The result follow from applying the same property of positive semi–definite matrices once more for $\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta$.

Finally, concerning 4, we need to study the difference

$$\begin{aligned}
&\Sigma_{R^eR^e} - \left(\Sigma_{R^eR^e} - \left(1-\lambda'(\beta'\Sigma_{R^eR^e}^{-1}\beta)\lambda\right)\left[\Sigma_{R^eR^e} - \beta(\beta'\Sigma_{R^eR^e}^{-1}\beta)^{-1}\beta'\right]\right) \qquad (1.A.21)\\
&= \left(1-\lambda'(\beta'\Sigma_{R^eR^e}^{-1}\beta)\lambda\right)\left[\Sigma_{R^eR^e} - \beta(\beta'\Sigma_{R^eR^e}^{-1}\beta)^{-1}\beta'\right]
\end{aligned}$$

Positive definiteness of $\left[\Sigma_{R^eR^e} - \beta(\beta'\Sigma_{R^eR^e}^{-1}\beta)^{-1}\beta'\right]$ follows from Lemma 1.13. ■

Proof of Theorem 1.6. Note that $\hat{\beta}$ converges to β and $\hat{\alpha}$ converges to α in probability.

$$\begin{aligned}
\beta &= \mathrm{Cov}\left[R^e, F^T\right]\Sigma_{FF}{}^{-1}, \qquad (1.A.22)\\
&= \mathrm{Cov}\left[R^e = \alpha^* + \beta^*F + \delta^*G + \varepsilon^*, F^T\right]\Sigma_{FF}{}^{-1},\\
&= \beta^* + \delta^*\mathrm{Cov}\left[G, F^T\right]\Sigma_{FF}^{-1}.
\end{aligned}$$

Now, note that

$$\begin{aligned}
\alpha &= \mathrm{E}\left[R^e\right] - \beta\mathrm{E}\left[F\right], \qquad (1.A.23)\\
&= \alpha^* + \beta^*\mathrm{E}\left[F\right] + \delta^*\mathrm{E}\left[G\right] - \beta\mathrm{E}\left[F\right],\\
&= \alpha^* + (\beta^* - \beta)\mathrm{E}\left[F\right] + \delta^*\mathrm{E}\left[G\right].
\end{aligned}$$

Furthermore, for $\hat{\lambda}$, first notice that

$$\hat{\lambda} = \left(\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\beta}\right)^{-1}\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathrm{E}\left[R^e\right] \qquad (1.A.24)$$

Then, observe following equality

$$\bar{R}^e = \hat{\beta}\lambda_F + \left(\bar{R}^e - \mathrm{E}\left[R^e\right]\right) - (\hat{\beta} - \beta)\lambda_F + (\beta^* - \beta)\lambda_F + \delta^*\lambda_G. \tag{1.A.25}$$

Multiplying both sides by $\left(\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\beta}\right)^{-1}\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ gives

$$\begin{aligned} \hat{\lambda} = \lambda_F &+ \left(\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\beta}\right)^{-1}\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\left[\left(\bar{R}^e - \mathrm{E}\left[R^e\right]\right) - (\hat{\beta} - \beta)\lambda_F\right] \\ &+ \left(\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\beta}\right)^{-1}\hat{\beta}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\left[(\beta^* - \beta)\lambda_F + \delta^*\lambda_G\right]. \end{aligned} \tag{1.A.26}$$

Hence, the probability limit of $\hat{\lambda}$ from GMM (1.3.1) is given by

$$\lambda_F + \left(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta\right)^{-1}\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\left[(\beta^* - \beta)\lambda_F + \delta^*\lambda_G\right] \tag{1.A.27}$$

■

Proof of Theorem 1.7. Note that $\hat{\beta}$ converges to β and $\hat{\alpha}$ converges to α in probability.

$$\begin{aligned} \beta &= \mathrm{Cov}\left[R^e, F^T\right]\Sigma_{FF}{}^{-1}, \\ &= \mathrm{Cov}\left[R^e = \alpha^* + \beta^*F + \delta^*G + \varepsilon^*, F^T\right]\Sigma_{FF}{}^{-1}, \\ &= \beta^* + \delta^*\mathrm{Cov}\left[G, F^T\right]\Sigma_{FF}^{-1}. \end{aligned} \tag{1.A.28}$$

Now, note that

$$\begin{aligned} \alpha &= \mathrm{E}\left[R^e\right] - \beta\mathrm{E}\left[F\right], \\ &= \beta^*\lambda_F + \delta^*\lambda_G - \beta\lambda_F, \\ &= (\beta^* - \beta)\lambda_F + \delta^*\lambda_G. \end{aligned} \tag{1.A.29}$$

Furthermore, for $\hat{\lambda_F}$, notice that $\hat{\lambda_F} = \bar{F}$, which converges to $\lambda_F = \mathrm{E}\left[F\right]$ in probability.

■

Proof of Corollary 1.3. Proof of the first part of the corollary: Note that

$$\beta\lambda_F = \beta\left(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta\right)^{-1}\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\mathrm{E}\left[R^e\right] \tag{1.A.30}$$

Note that $\hat{\beta}\hat{\lambda}$ is consistent for $\mathrm{E}\left[R^e\right]$ if and only if $\mathrm{E}\left[R^e\right] = \beta\left(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta\right)^{-1}\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\mathrm{E}\left[R^e\right]$ which is equivalent to

$$\left[I_N - \beta\left(\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\beta\right)^{-1}\beta'\Sigma_{\varepsilon\varepsilon}^{-1}\right]\mathrm{E}\left[R^e\right] = 0 \tag{1.A.31}$$

To prove the second part of the corollary, note that $\hat{\beta}\ hat\lambda$ converges to $\beta\lambda$. Using 1.A.28 and $\lambda_F = \mathrm{E}\left[F\right]$ (based on 1.7), we have

$$\begin{aligned} \beta\lambda_F &= (\beta^* + \delta^*\mathrm{Cov}\left[G, F^T\right]\Sigma_{FF}^{-1})\lambda_F, \\ &= \mathrm{E}\left[R^e\right] - ((\beta^* - \beta)\lambda_F + \delta^*\lambda_G). \end{aligned} \tag{1.A.32}$$

■

Proof of Theorem 1.8. Consistency of $\hat{\alpha} + \beta\lambda_F$ is straightforward. The asymptotic variance is given by the delta method for the function g. Assume $g(\alpha, \beta, \lambda_F) = \alpha + \beta\lambda_F$. The asymptotic covariance matrix of $\alpha, \beta and \gamma$ is given in Lemma 1.11 (denoted by V). We have, by applying the delta method, that

$$\sqrt{T}\left(g(\hat{\alpha}, \hat{\beta}, \hat{\lambda}) - g(\alpha, \beta, \lambda)\right) \xrightarrow{d} \mathcal{N}(0, \dot{g}'V_{\alpha,\beta,\lambda}\dot{g}), \tag{1.A.33}$$

with

$$\dot{g} = \begin{bmatrix} [1 \quad \lambda'] \otimes I_N & \beta \end{bmatrix}.$$

Matrix multiplication of calculating $\dot{g}'V_{\alpha,\beta,\lambda}\dot{g}$ gives $\Sigma_{R^eR^e}$.

■

Chapter 2

Efficient Estimation of Integrated Volatility and Related Processes

Abstract

We derive nonparametric bounds for inference about functionals of high-frequency volatility, in particular, integrated power variance. In the absence of microstructure noise, we find that standard realized variance attains the nonparametric efficiency bound, also in case of unequally spaced random observation times. For higher powers, e.g., integrated quarticity, the block-based procedures of Mykland and Zhang (2009) can get arbitrarily close to the nonparametric bounds in case of equally spaced observations. The estimator in Jacod and Rosenbaum (2013) is efficient, also at non-constant volatility, still for equally spaced data. For unequally spaced data, we provide an estimator, similar to that of Kristensen (2010), that can get arbitrarily close to the nonparametric bound. Finally, contrary to public opinion, we demonstrate that parametric information about the functional form of volatility generally leads to a decreased lower bound, unless the volatility process is piecewise constant.

2.1 Introduction

The availability of high-frequency data has led to the development of new estimators of integrated volatility and their asymptotic properties (see Andersen and Bollerslev (1998), Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002a,b), among many others). Inference from high-frequency data is not only confined to integrated volatility. There is a strand of literature on estimating power variances, or other smooth transformations, based on the intraday price data (see Barndorff-Nielsen and Shephard (2003, 2004), Jacod (2008), Mykland and Zhang (2009), Kristensen (2010), and Jacod and Rosenbaum (2013) to name just a few).

We focus in this paper on the issue of *efficient* estimation of integrated smooth transformations of instantaneous variances from two perspectives. First, we analyze the efficiency of estimators proposed in the literature and, actually, propose a new estimator that can deal with unequally-spaced (random, but predictable) observation times. Secondly, from the perspective of modeling, we detail the efficiency gains possible when a researcher is willing to make parametric assumptions on the volatility path or, equivalently, we characterize which volatility specifications are adaptive. For this, we analyze, in a concrete probabilistic setting, limiting experiments concerning inference about integrated functions of volatility. To be precise, for the derivation of the efficiency bounds, we consider returns, conditionally to the realizations of both the volatility function and the sampling times, to be normally distributed. This setting is deliberately much simpler than the assumptions that are usually imposed in the literature about realized quantities that often consider general Itô semimartingales, possibly contaminated by some micro structure noise.

Our work complements two recent papers in various ways. First, Clément et al. (2013) derive a locally asymptotically mixed normal (LAMN) limiting experiment, assuming that volatility follows a diffusion process. Their Proposition 2 is closely related to our Theorem 2.3, though conceptually different. To be precise, we consider the pathwise properties of volatility such that the model is locally asymptotically normal (LAN), i.e., we consider the inference problem conditionally on the realized intraday path of volatility. This does not mean that we assume volatility to be deterministic, but the statistical experiments we consider are conditional on the realization of the volatility path. This approach is also taken in Reiß (2011). We precisely identify the pathwise properties needed for LAN and show that, for instance, jumps in volatility are not excluded. Moreover, we don't need to assume that the volatility process is a semimartingale. It might also entail, for instance, some fractional Brownian motion component to accommodate long memory in volatility (see, e.g., Comte and Renault (1998)). Although in their more abstract results, Clément et al. (2013) allow for (deterministic) irregularly spaced observation times, their discussion about the efficiency of existing estimators of integrated power variance focuses on the case of regularly spaced data only. We consider even random, albeit predictable, irregularly spaced observation times. Moreover, we also provide a new (nearly) efficient estimator in this case. Second, our paper complements Reiß (2011) as that paper focuses on the limiting experiments arising when prices are contaminated with (market micro structure) noise. This turns out to lead to fundamentally different limiting experiments and, even, to different optimal rates of convergence. In the absence of noise, Reiß (2011)'s limit experiments are no longer valid. Moreover, the absence of noise leads to less restrictive assumptions on the sample paths of volatility, in particular, allowing jumps. We discuss these links in more detail in the remainder of the paper.

The local asymptotic normality result we derive in our simple setting (which allows us to explicitly analyze likelihood ratio processes of a fairly simple form) leads to a well-defined optimality concept for asymptotic inference using the so-called convolution theorem. As stated in Jacod and Rosenbaum (2013), "... even for the simpler problem of estimating integrated volatility, the concept of efficiency in the general nonparametric or semi-parametric setting is not well established so far".

Accordingly, they decide to generally call "efficient" a procedure which is efficient in the usual sense for the submodel assuming a time-invariant volatility, i.e., what they call a "toy" model where observations are generated by a constant volatility Brownian motion. However, this simple concept of efficiency is not sufficient for at least two reasons. First, if a nonparametric estimator attains a bound induced by a parametric submodel, the nonparametric estimator can indeed be called efficient, but only for data generating processes that belong to this parametric submodel. One of the consequences of our paper is that the Jacod and Rosenbaum (2013) estimator is even nonparametrically efficient for non-constant volatility paths, at least in case of regularly spaced data. Second, indeed, if the observation times are irregular, like transactions times or times of quote changes, then these simple parametric submodels may be misleading about nonparametric efficiency. Hayashi et al. (2011) give a counterexample considering the estimation of a submodel with time-invariant volatility with irregular sampling times. For this simple model, the maximum likelihood estimator (MLE) is asymptotically efficient and easy to compute as an average of squared returns divided by corresponding durations. Unfortunately, this MLE formula does not even deliver a consistent estimator of integrated variance if volatility is time varying and the sampling times are irregular. In other words, the quasi-maximum likelihood estimation approach of Xiu (2010) with regularly sampled high-frequency data, based on a quasi-likelihood method as if the volatility were constant, does not work with irregular sampling. Our paper generalizes the counterexample of Hayashi et al. (2011) by showing that even in case of a parametric model defined by a piecewise constant volatility, the parametric efficiency bound for estimating integrated powers of the volatility does not coincide with the nonparametric bound. It takes an ad hoc assumption (namely, the specific power of volatility properly rescaled according to the density of observation times being piecewise constant) in order to obtain equality of the parametric and nonparametric efficiency bounds.

The present paper offers three important contributions. First or all, we extend the analysis of efficiency of estimators for integrated transformations of instantaneous variance to the situation of irregularly spaced, random but predictable, sampling times. It follows that, for integrated variance, realized variance remains nonparametrically efficient in this case. Our results also show how the denseness of observations throughout the day affects the possible precision of estimators. We provide an estimator, similar to the one proposed in Kristensen (2010), that is nearly efficient also in case observations times are irregular. The "near" efficiency signifies that the limiting variance of our estimator can get arbitrarily close to the nonparametric lower bound, just like the block-based procedure in Mykland and Zhang (2009). We have not yet been able to derive, for the case of irregularly spaced data, a fully efficient estimator like in Jacod and Rosenbaum (2013) for the regularly spaced situation.

Second, we detail the pathwise properties of volatility needed to obtain our LAN result. This is summarized in the new concept of (sample paths of) locally bounded variance, a concept that does not rule out jumps in volatility and is satisfied, e.g., by the sample paths of Brownian motion. We expect, though were unable to prove formally, that this condition is much more generally satisfied by sample paths of (Brownian) semimartingales.

Third, we show for which volatility paths and parametric volatility specifications, the nonparametric and parametric lower bound coincide, i.e., when the nonparametric model is actually adaptive. This is useful if in empirical work one is willing to take misspecification risk in return for efficiency gains. We show that these gains indeed can be sizable. This has sometimes been overlooked in the literature, in particular due to the fact that realized variance achieves the parametric lower bound for constant volatility specifications (with and without regularly spaced observations).

The rest of paper is organized as follows. Section 2.2 introduces our model setup and provides the local properties of volatility paths which are assumed for our analysis. In particular, the functional parameter of interest is made more explicit to address the efficiency issue and the distribution of possibly random observation times is discussed. Section 2.2 states the set of maintained assumptions about both the return process and the sampling scheme sufficient to derive our asymptotic results. In Section 2.3 we obtain our lower bounds for (smooth) functionals of volatility and show, for equidistant data, when the nonparametric efficiency bound is attained by existing estimators as those proposed by Mykland and Zhang (2009) and Jacod and Rosenbaum (2013). In Section 2.4, we discuss a (nearly) efficient estimator for integrated smooth functions of volatility, also in case observations are irregularly spaced in time. Subsequently, Section 2.5 shows in which circumstances parametric volatility models lead to additional information that can be exploited statistically, i.e., when MLE improves upon the nonparametric efficiency bound. Finally, Section 2.6 concludes and the appendix gathers some proofs.

2.2 Setting and pathwise properties

We are interested in the pathwise properties of a univariate (instantaneous variance) process $\sigma^2 = \{\sigma^2(t) : 0 \leq t \leq 1\}$. All processes are assumed to be adapted to a (given) filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\leq t\leq 1}, \mathbb{P}\right)$. Inference about (the paths of) σ^2 will be based on observations on a, say, log-price process S at random sampling times $t_{i,n}$. Section 2.2.1 introduces the properties required on these sampling times, while Section 2.2.2 discusses further assumptions on S. Section 2.2.3 introduces a concept related to the Quadratic Variation of Time by Mykland and Zhang (2006) that we will need to clarify the impact of the denseness of observations on efficiency bounds.

2.2.1 The sampling scheme

At stage n, we consider a strictly increasing sequence of stopping times $t_{i,n}$, with $i = 0, 1, \ldots, N_n$ and $0 = t_{0,n} < t_{1,n} < \ldots < t_{N_n,n} \leq 1$. The deterministic sequence $n = 1, 2, \ldots$ is used to index the experiments whose limit we will consider. Note that N_n stands for the actual number of observations available at stage n, and this number may be random. Also, even though the volatility process σ^2 is never observed directly, we call, for now with some abuse, the times $t_{i,n}$ "observation times". In addition, the double array $(t_{i,n})_{0\leq i\leq N_n; n\geq 1}$ of stopping times forms the "sampling scheme".

We assume the mesh of the sampling scheme to converge to zero at some deterministic rate and N_n/n to be bounded from above, almost surely. As the sequence n is an index to define the asymptotic setup, we can always choose N_n to be bounded by n. A standard assumption would be to assume the mesh of the sampling scheme to converge to zero at rate n^{-1}. However, we relax this assumption in order to include, for instance, Poisson sampling with intensity of the order $O(n)$ for which N_n/n converges to 1 but the mesh can only be bounded by $O(\log(n)/n)$.

Assumption 2.1. *We suppose that the sampling scheme is a double array $(t_{i,n})_{0\leq i\leq N_n; n\geq 1}$ of stopping times with respect to $\{\mathcal{F}_t\}_{0\leq t\leq 1}$ such that*

- $0 = t_{0,n} < t_{1,n} < \ldots < t_{N_n,n} \leq 1$;
- *for all n, $N_n \leq n$ a.s.;*
- $\sqrt{n}\max|t_{i,n} - t_{i-1,n}| = o_P(1)$, *as $n \to \infty$;*
- $n\sum_{i=1}^{n}(t_{i,n} - t_{i-1,n})^2 = O_P(1)$ *as $n \to \infty$.*

Moreover, we maintain the (restrictive) assumption that the stopping times are strongly predictable. Obviously, this does cover the case of irregularly, but deterministically, spaced observation times.

Assumption 2.2. *The sampling scheme* $(t_{i,n})_{0\leq i\leq N_n; n\geq 1}$ *is* $\{\mathcal{F}_t\}_{0\leq t\leq 1}$*-predictable; that is* $t_{i,n}$ *is* $\mathcal{F}_{t_{i-1,n}}$ *measurable, for all* i, n.

It is worth acknowledging that in the spirit of Assumption (C) in Hayashi et al. (2011), it would be possible to relax Assumption 2.2 so that, conditionally on $\mathcal{F}_{t_{i-1,n}}$, $t_{i,n}$ is independent of the processes of interest. Allowing for genuinely endogenous sampling times, albeit realistic for transaction times, much complicates the specification of the conditional distribution of returns given the observation times. As extensively discussed by Li et al. (2014), it significantly changes the asymptotic distribution of, e.g., realized volatility. A study of nonparametric efficiency in this more general setting appears to be a daunting task, beyond the scope of this paper.

The pathwise behavior of the volatility process σ of interest has to be restricted by a regularity condition. To formalize this, we introduce the concept of locally bounded variance.

Definition 2.1. *Let* f *be a real-valued cadlag function on the interval* $[0, 1]$. *We say that* f *is of* locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance *if, for any sampling scheme* $(t_{i,n})_{0\leq i\leq N_n; n\geq 1}$ *satisfying Assumption 2.1 and 2.2,*

$$\max_{n\in\mathbb{N}} \left\{ \sum_{i=1}^{N_n} V\left(f|_{t_{i-1,n}}^{t_{i,n}} \right) \right\} \tag{2.2.1}$$

is almost surely bounded, where

$$V\left(f|_{t_{i-1,n}}^{t_{i,n}} \right) = \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} \left(f(u) - \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} f(v)dv \right)^2 du. \tag{2.2.2}$$

Throughout the paper, we assume that the sample paths of σ^2 are of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance. Observe that Lemma 2.6 and the observation that $\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}} \right)$ equals $\mathrm{V}\left(f(U)\right)$ when U is uniformly distributed over $(t_{i-1,n}, t_{i,n}]$, implies that then also the sample paths of σ^{-2} are of locally bounded variance. As this assumption is key to the asymptotic theory developed in the next sections, it is worth analyzing whether the condition holds for paths of often-used volatility processes.

First of all, note that if f and g are of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance, so is their sum and are scalar products. The functions of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance thus form a vector space. Second, we have the bound, for $0 \leq i \leq N_n$,

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) \leq \frac{1}{4}\left[\max_{t_{i-1,n}\leq u\leq t_{i,n}} f(u) - \min_{t_{i-1,n}\leq u\leq t_{i,n}} f(u)\right]^2, \tag{2.2.3}$$

so that

$$\sum_{i=1}^{N_n}\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) \leq \frac{1}{4}\sum_{i=1}^{N_n}\left[\max_{t_{i-1,n}\leq u\leq t_{i,n}} f(u) - \min_{t_{i-1,n}\leq u\leq t_{i,n}} f(u)\right]^2. \tag{2.2.4}$$

As a result, monotonic functions f are of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance and, hence, so are functions of finite variation. In particular, our analysis does not rule out jumps in volatility. For functions not of finite variation, the analysis is more subtle but diffusions are not excluded; see Appendix 2.A for details.

2.2.2 Functional parameter and model

Inference about (functionals of) the path of σ^2 will be based on observations of a process $S = \{S(t) : 0 \leq t \leq 1\}$, which can be thought of as a log-price process, at the sampling times $(t_{i,n})_{0\leq i\leq N_n; n\geq 1}$. We throughout assume that S is $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-adapted. In this section, we formalize the underlying data generating mechanism for the derivation of the efficiency bound. Essentially, we consider returns, conditionally on both the realizations of the volatility and the sampling times, to be normally distributed. We stress that our kernel-based estimator in Section 2.4 does not require Gaussianity of the returns.

Assumption 2.3. *Suppose that* $R_{i,n} := \log\left\{S_{t_{i,n}}/S_{t_{i-1,n}}\right\}$ *is the log-return over the* *time interval* $(t_{i-1,n}, t_{i,n}]$. *Then, conditionally on* $\mathcal{F}_{t_{i-1,n}}$, $R_{i,n}$ *is distributed as*

$$N\left(\left[\mu(t_{i-1,n}, t_{i,n}) + \gamma(t_{i-1,n}, t_{i,n})\sigma^2(t_{i-1,n}, t_{i,n})\right]\Delta t_{i,n}, \sigma^2(t_{i-1,n}, t_{i,n})\Delta t_{i,n}\right), \quad (2.2.5)$$

where $\mu(t_{i-1,n}, t_{i,n})$, $\sigma^2(t_{i-1,n}, t_{i,n})$ *and* $\gamma(t_{i-1,n}, t_{i,n})$ *are bounded and* $\mathcal{F}_{t_{i-1,n}}$*-measurable.*

It's worth stressing that for our LAN result it is not needed that *investors* know quantities as $\mu(t_{i-1,n}, t_{i,n})$, $\sigma^2(t_{i-1,n}, t_{i,n})$, or $\Delta t_{i,n}$ at time $t_{i-1,n}$. They may see the volatility (and the subsequent times of trades) as stochastic and their conditional variance at time $t_{i-1,n}$ of the return $R_{i,n}$ can be computed as the projection of the above moments on the σ-field subset of $\mathcal{F}_{t_{i-1,n}}$ that describes the investors' information at time $t_{i-1,n}$. The sequence of σ-fields $\mathcal{F}_{t_{i-1,n}}$, $i = 1, \ldots, N_n$ is a modeling tool for the sake of specifying a statistical model and should not necessarily be interpreted in terms of investors' information. Note that, precisely contrary to the limiting approximations discussed in Reiß (2011), we exclude possible (micro structure) contamination of observed returns. As explained in the introduction, the resulting limiting experiments are materially different.

We assume that the information brought by asset returns is exogenous in the following sense.

Assumption 2.4. *The filtration* $\{\mathcal{F}_t\}_{0\le t\le 1}$ *is the natural one of* $\{(S(t), Z(t)) :\ 0 \le t \le 1\}$, *where* $\{Z(t) : 0 \le t \le 1\}$ *is a (possibly multivariate) stochastic process of exogenous variables such that, given* $\mathcal{F}_0$ *and* N_n, *the joint conditional density of* $\{(S(t), Z(t)) :\ 0 \le t \le 1\}$ *can be written*

$$p\left(\left\{Z_{t_{i,n}}\right\}_{1\le i\le N_n}\right)\prod_{i=1}^{N_n} p\left(R_{i,n}|\mathcal{F}_{t_{i-1,n}}\right). \quad (2.2.6)$$

Remark 2.1. Assumption 2.4 requires non-causality (in the Sims sense) from the return process $(R_{i,n})_{1\le i\le N_n}$ to the $(Z_{t_{i,n}})_{1\le i\le N_n}$. Otherwise, the conditioning information in $p(R_{i,n}|\mathcal{F}_{t_{i-1,n}})$ should also involve future values $Z_{t_{j,n}}$, $j > i$. We also preclude instantaneous causality in order to erase the contemporaneous value $Z_{t_{i,n}}$ in the conditioning information that defines the probability distribution of $R_{i,n}$. Up to discussions regarding initial values, Sims' non-causality is known to be equivalent to Granger non-causality. In other words, we basically assume that returns do not Granger-cause state variables, like stochastic volatility.

It is generally convenient to study the return variance as an integral over the corresponding event interval of the so-called spot volatility process $\sigma^2 = \{\sigma(t) : 0 \leq t \leq 1\}$, so that:

$$\sigma^2(t_{i-1,n}, t_{i,n})\,[t_{i,n} - t_{i-1,n}] = \int_{t_{i-1,n}}^{t_{i,n}} \sigma^2(u)\mathrm{d}u. \tag{2.2.7}$$

In other words, $\sigma^2(t_{i-1,n}, t_{i,n})$ is the arithmetic mean of the function σ^2 over the interval $(t_{i-1,n}, t_{i,n}]$. This situation occurs, for instance when the returns are generated from log-prices S_t satisfying the differential equation

$$\mathrm{d}\log S_t = a_t\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{2.2.8}$$

for some appropriate drift a_t and Brownian motion W_t.

We introduce the following notations and assumption. Let $D[0,1]$ denote the set of real-valued cadlag functions on the interval $[0,1]$ and let $D_+[0,1]$ be the subset of functions that take strictly positive values only. Both spaces are equipped with the supremum norm $\|\cdot\|$. Assumption 2.5 will not be imposed in the limiting results of our kernel-based estimator in Section 2.4.

Assumption 2.5. *The parameter space Ξ is the set of all elements σ^2 in $D_+[0,1]$ such that σ^2 is bounded away from zero and of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance.*

2.2.3 On the denseness of events

In order to analyze the consequences of irregularly spaced event times, we introduce a process representing effectively the time change induced by the observation times $t_{i,n}$. We define

$$T_n(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_{i,n}-t_{i-1,n}}\left[u \wedge t_{i,n} - u \wedge t_{i-1,n}\right], \qquad 0 \leq u \leq 1. \tag{2.2.9}$$

Observe that the function T_n is piecewise linear and increases over each time interval $(t_{i-1,n}, t_{i,n}]$ exactly an amount $1/n$. As a result $T_n(t_{i,n}) = i/n$, $i = 0, \ldots, n$. Thus, at the observation times $u = t_{i,n}$, $T_n(u)$ coincides with the empirical distribution function of these observation times. For regularly spaced data $t_{i,n} = i/n$ and $T_n(u) = u$. The function T_n is also closely related to what Mykland and Zhang (2006) call the Quadratic Variation of Time; a relation that we will make precise below. Also, note that the sum in (2.2.9) is till $i = n$, and not, as before, till $i = N_n$. This is an abuse of notation that we will maintain throughout the remainder of the paper and is warranted in view of (2.A.3).

We impose the following additional assumption on the observation times $t_{i,n}$.

Assumption 2.6. *The function T_n as defined in (2.2.9) converges almost surely to a distribution function T on $[0,1]$ in the topology of weak convergence. That is, for any function $f \in D[0,1]$,*

$$\int_{u=0}^{1} f(u)\,dT_n(u) \to \int_{u=0}^{1} f(u)\,dT(u),$$

almost surely, as $n \to \infty$. Moreover, T admits a strictly positive and bounded density T' and $T_n' \to T$, almost surely.

Remark 2.2. Assumption 2.6 actually has a few strong consequences. As the limit T is continuous, the functions T_n monotone, and $[0,1]$ compact, the convergence of T_n to T is actually uniform (see Buchanan and Hildebrandt (1908)). Consequently, again using the continuity of T, the quantile functions T_n^{-1} also converge weakly, pointwise, and, by the same argument, uniformly to T^{-1}.

It's informative to relate Assumption 2.6 to the concept of Asymptotic Quadratic Variation of Time (AQVT) as in Mykland and Zhang (2006). Observe

$$T_n'(u) = [n(t_{i,n} - t_{i-1,n})]^{-1} \text{ for } t_{i-1,n} < u < t_{i,n}.$$

Consequently, for $u \in [0,1]$,

$$H_n(u) \quad := \quad \int_0^u \frac{1}{T_n'(z)} \mathrm{d}z = n \sum_{t_{i,n} \le u} (t_{i,n} - t_{i-1,n})^2 + n(t_{i^*+1,n} - t_{i^*,n})(u - t_{i^*,n}),$$

where $t_{i^*,n} \le u \le t_{i^*+1,n}$. In general convergence of T_n and H_n cannot be related, but, under appropriate additional smoothness, one would have

$$H(u) := \lim_{n\to\infty} H_n(u) = \lim_{n\to\infty} \int_0^u \frac{1}{T_n'(z)} \mathrm{d}z, \tag{2.2.10}$$

and thus $H'(u) = 1/T'(u)$. We represent the denseness of events in terms of T_n and T, rather than H_n and H, as this notion arises naturally in the study of the likelihood ratios in the next section. Following Mykland and Zhang (2012), it is easy to check that $H'(u) \equiv 1$ (or, equivalently, $T'(u) \equiv 1$) if and only if

$$\sum_{i=1}^{N_n} \left(t_{i,n} - t_{i-1,n} - \frac{1}{n} \right)^2 = o_P(1).$$

We call such a sampling scheme (asymptotically) regular.

2.3 Lower bounds for integrated functions of variance

As indicated in the introduction, we base our optimality criteria on the Hájek-Le Cam theory of convergence of experiments. We refer to, e.g., van der Vaart

(2000) for details. We actually show in this section that appropriately parametrized local versions of the model described in Assumptions 2.1-2.6 converge, as n tends to infinity, to a Gaussian shift experiment, i.e., our experiment is locally asymptotically normal (see van der Vaart (2000), Section 7). Using by now standard arguments, the least-favorable of these parametric submodels describes the lower bound for estimating functionals of the volatility path, see Section 2.3.

As is to be expected, realized variance plays a key role in the analysis. The idea of estimating volatility of returns over a fixed interval as the sum of squared realizations given the availability of sufficiently high sampling frequency was noted already in Merton (1980). More recently, realized variance measures constructed from intraday data have been exploited by Taylor and Xu (1997) and Andersen, Bollerslev, and Lange (1999), among others. We define the realized variance process as

$$RV_n(u) = \sum_{i=1}^{n} R_{i,n}^2 I\left\{t_{i,n} \leq u\right\}, \qquad 0 \leq u \leq 1. \tag{2.3.1}$$

Observe that $RV_n(1)$ coincides with the standard definition of realized variance, but we need this process version below to clarify the role of unequally spaced time points $t_{i,n}$ later. It turns out that in the local asymptotic normality result another process plays an important role. We define the duration-weighted realized variance process as

$$RV_n^*(u) = \frac{1}{n}\sum_{i=1}^{n} \frac{R_{i,n}^2}{\left(t_{i,n} - t_{i-1,n}\right)^2}\left[u \wedge t_{i,n} - u \wedge t_{i-1,n}\right], \qquad 0 \leq u \leq 1. \tag{2.3.2}$$

Note that RV_n^* is a piecewise linear process that increases by an amount of $R_{i,n}^2/\left[n\left(t_{i,n} - t_{i-1,n}\right)\right]$ over the interval $(t_{i-1,n}, t_{i,n}]$. We could have used a piecewise constant definition, similar to the definition of the realized variance process RV, but the continuity induced by the linear interpolation turns out to be mathematically convenient.

We first provide a joint functional central limit result for the realized variance

RV_n and the duration-weighted realized variance RV_n^* above. A proof is again provided in the appendix. As our nearly efficient estimator in Section 2.4 takes this result as input, we formulate it here as a condition. Thus, once more, our estimator is valid under much more general conditions than Assumptions 2.1-2.6. In particular, it does not rely on returns being normally distributed.

Condition 1. *The processes*

$$\sqrt{n}\left[RV_n(u) - \int_{v=0}^{u} \sigma^2(v)\,dv\right] \tag{2.3.3}$$

and

$$\sqrt{n}\left[RV_n^*(u) - \int_{v=0}^{u} \sigma^2(v)\,dT_n(v)\right] \tag{2.3.4}$$

jointly converge to $\sqrt{2}\int_{v=0}^{u} \sigma^2(v)T'(v)^{-1/2}dW(v)$ *and* $\sqrt{2}\int_{v=0}^{u} \sigma^2(v)T'(v)^{+1/2}dW(v)$*, respectively, where* W *denotes a standard Brownian motion. The convergence is weakly in* $(D[0,1], \|\cdot\|)$.

Lemma 2.1. *Under Assumptions 2.1-2.6, Condition 1 holds.*

As mentioned before, the nonparametric analysis where σ^2 denotes a functional parameter is, in line with classical reasoning, reduced by considering parametric submodels and, then (see Theorem 2.3), considering the least-favorable among them. Thus, fix $\sigma_0^2 \in D_+[0,1]$ and define local alternatives for $h \in D[0,1]$, with $\|h\| \leq 1$, by

$$\sigma_{\alpha/\sqrt{n}}^{-2}(u) = \sigma_0^{-2}(u)\left[1 + \frac{\alpha}{\sqrt{n}}h(u)\right], \qquad \alpha \in (-1,1). \tag{2.3.5}$$

We write shorthand $\mathbb{P}_\alpha^{(n)}$ for the probability measure induced by $R_{1n}, \ldots, R_{nn}$ under $\sigma_{\alpha/\sqrt{n}}^2$, i.e., $\mathbb{P}_\alpha^{(n)} = \mathbb{P}_{\sigma_{\alpha/\sqrt{n}}^2}^{(n)}$.

Theorem 2.2. *Under Assumptions 2.1-2.6, the experiment* $\left\{\mathbb{P}_\alpha^{(n)} : \alpha \in (-1,1)\right\}$ *is* *asymptotically normal with*

$$\log \frac{\mathrm{d}\mathbb{P}_\alpha^{(n)}}{\mathrm{d}\mathbb{P}_0} = -\frac{\alpha}{2}\int_0^1 \sigma_0^{-2}(u)h(u)d\sqrt{n}\left[RV_n^*(u) - \int_{v=0}^u \sigma_0^2(v)dT_n(v)\right] -\frac{\alpha^2}{4}\int_{u=0}^1 h^2(u)dT(u). \tag{2.3.6}$$

The Fisher Information is given by $\frac{1}{2}\int_{u=0}^1 h^2(u)dT(u)$. *In particular, the probability measures* $\mathbb{P}_\alpha^{(n)}$ *and* $\mathbb{P}_0^{(n)}$ *are contiguous.*

For more details on local asymptotic normality results as in Theorem 2.2, we refer the reader to van der Vaart (2000). In the following section, we will use it to establish lower bounds on the precision of regular estimators for integrals of smooth transformations of instantaneous variances, in the absence of micro structure noise.

Assume now that we are interested in estimating the following generalized version of the standard integrated volatility

$$\psi_g\left(\sigma^2\right) = \int_0^1 g(u, \sigma^2(u))\omega(u)\mathrm{d}u, \tag{2.3.7}$$

where ω is a known weighting function on $[0,1]$ and $g : [0,1] \times \mathbb{R} \to \mathbb{R}$ is a a known time-dependent transformation. This setup includes most standard measures.

(i) For $\omega(u) \equiv 1$ and $g(u, \sigma^2(u)) = \sigma^{2p}(u)$ we have the so-called "power variation" as studied in particular by Jacod (2008). These power variations are popular in particular to assess the asymptotic variance of estimators of power variation of lower order. For instance, quarticity ($p = 2$) is informative about the asymptotic variance of realized variance ($p = 1$). A non-flat weighting function ω may be necessary to accomodate the effect of irregular sampling.

(ii) For $\omega(u) \equiv 1$ and $g(u, \sigma^2(u)) = \exp\left(-s\sigma^2(u)\right)$ for some given $s \in \mathbb{R}_+$, we have the empirical Laplace transform function of the volatility process. A consistent asymptotically (mixed) normal estimator of the Laplace transform has been provided by Todorov and Tauchen (2012). Li, Tauchen and Todorov (2013) subsequently use this estimator to estimate the volatilty occupation time corresponding to $g(u, \sigma^2(u)) = 1_{]0,x]}(\sigma^2(u))$ for some given $x \in \mathbb{R}_+$. However, this latter example will not be covered here since, in order to compute Cramér-Rao efficiency bounds, we always assume that $g(u, \sigma^2)$ is a continuously differentiable function of the underlying spot variance σ^2.

(iii) Time-dependent transformations of volatility $g(u, \sigma^2(u))$ may be relevant when computing implied volatilities from option prices. These option prices typically depend not only on the underlying spot volatility but also on the time to maturity.

Along the paths induced by $\sigma^2_{\alpha/\sqrt{n}}$ defined in (2.3.5) we have

$$\begin{aligned}
\psi_g\left(\sigma^2_{\alpha/\sqrt{n}}\right) &= \int_0^1 g\left(u, \sigma_0^2(u)\left(1+\frac{\alpha}{\sqrt{n}}h(u)\right)^{-1}\right)\omega(u)\mathrm{d}u \\
&= \psi_g\left(\sigma_0^2\right) - \frac{\alpha}{\sqrt{n}}\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u + o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

as $n \to \infty$. This expansion is valid uniformly in h for $\|h\| \leq 1$ if g fulfills the following assumption.

Assumption 2.7. *$g(u, \sigma^2)$ is continuous in u and continuously differentiable in σ^2.*

As a result of the above expansion, the Fréchet derivative of our parameter of interest $\psi_g\left(\sigma^2\right)$ with respect to $\alpha/\sqrt{n})$ is given by

$$-\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u.$$

We can now proceed as usual and derive the nonparametric lower bound for estimating $\psi_g\left(\sigma^2\right)$ as the largest bound obtained in the parametric models indexed by h.

More precisely, given h, we apply the Convolution Theorem as, for instance, stated in Bickel et al. (1993) Theorem 2.3.1. Consider a regular[7] estimator $\hat{\psi}_g^{(n)}$ for

[7] Regularity is needed to exclude pointwise superefficient estimators. Its definition does not necessarily require a Gaussian limiting distribution, but for our setting this situation suffices.

$\psi_g\left(\sigma^2\right)$ in the sense that, under $\mathbb{P}_\alpha^{(n)}$,

$$\sqrt{n}\left(\hat{\psi}_g^{(n)} - \psi_g\left(\sigma^2_{\alpha/\sqrt{n}}\right)\right) \to_L N(0, V). \tag{2.3.8}$$

Then, we know that V is at least equal to the squared derivative of $\psi_g\left(\sigma^2\right)$ (with respect to α) times the inverse of the Fisher Information, thus

$$V \geq \frac{\left(\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u\right)^2}{\int_0^1 h^2(u)dT(u)/2}. \tag{2.3.9}$$

We thus have the following nonparametric bound.

Theorem 2.3. *Under Assumptions 2.1-2.7, any regular estimator for $\psi_g\left(\sigma^2\right)$ as defined in (2.3.7) based on observations on the grid $t_{i,n}$, $i = 1, \ldots, n$, has, under $\mathbb{P}_{\sigma_0^2}^{(n)}$, a limiting variance of at least*

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\right)^2 \sigma_0^4(u)\frac{\omega^2(u)}{T'(u)}du \tag{2.3.10}$$

Proof. The least-favorable submodel is obtained by choosing h in the local alternatives (2.3.5) such that the parametric lower bound (2.3.9) is maximized. From Cauchy-Schwarz, we know that this happens for

$$h(u) \propto \frac{\frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)\omega(u)}{T'(u)}.$$

Pluging this least-favorable h into (2.3.9) gives the result. ∎

In order to discuss the relationship between Theorem 2.3 and the extant literature, it is worth to focus on the case of flat weights ($\omega(u) \equiv 1$) and to discuss two separate cases: first (asymptotically) regular sampling ($T'(u) \equiv 1$) and, second, general sampling schemes (arbitrary T').

2.3.1 Regular sampling without weighting

In case of regular sampling without weighting the efficiency bound (2.3.10) reduces to

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\right)^2 \sigma_0^4(u)\mathrm{d}u \tag{2.3.11}$$

This formula is a univariate version of formula (30) in Clément et al. (2013) for the case $g(u, \sigma^2) = g(\sigma^2)$. Recall, however, that Clément et al. (2013) derived this bound from a LAMN property, assuming that σ is generated by an Itô process, independent of the Brownian motion defining the return innovations. The validity of Theorem 2.3 is more general. If one wants to see σ as a stochastic process, independent of the leading Brownian motion, it can be any process whose sample paths are almost surely of locally bounded variance.

As far as the time-independent case $g(u, \sigma^2) = g(\sigma^2)$ is concerned, it is worth considering both examples above.

Example 1. Power variation: *The case of power variation is obtained using $g_p(\sigma^2) = \sigma^{2p}$ and leads to the nonparametric lower bound*

$$V(g_p) = 2p^2 \int_0^1 \sigma_0^{4p}(u)\,du. \tag{2.3.12}$$

Practical implications of this result have been known at least since Mykland and Zhang (2009). For $p = 1$, that is for the estimation of integrated variance, empirical quadratic variation is an (asymptotically) efficient estimator. Note that this case precisely corresponds to linear g and, hence, smoothing operations and the transformation g commute.

By contrast, for $p > 1$, the realized power variation does not deliver an efficient estimator. Indeed, Mykland and Zhang (2009) also study

$$\hat{\psi}_{g_p} = \frac{1}{E|N(0,1)|^{2p}} \sum_{i=1}^{n} (t_{i,n} - t_{i-1,n})^{1-p} \left|R_{i,n}^2\right|^p \tag{2.3.13}$$

and give, under somewhat different conditions, the limiting variance[8]

$$\frac{V|N(0,1)|^{2p}}{(E|N(0,1)|^{2p})^2}\int_{u=0}^{1}\sigma_0^{4p}(u)\,du. \tag{2.3.14}$$

But it can easily be shown that the coefficient in front of the integral exceeds $2p^2$ *if (and only if)* $p > 1$. *For instance, for* $p = 2$ *(integrated quarticity), the coefficient equals* 10.67, *while the lower bound is* $2p^2 = 8$*— an ARE of only* 75%. *This inefficiency is also noted in Jacod and Rosenbaum (2013).*

Mykland and Zhang (2009) also provide a block-based estimator for integrated powers of volatility whose limiting variance, for large block size (and still equally spaced data) is arbitrarily close to the efficiency bound (see their formula (63)). Even though Mykland and Zhang (2009) do not formally derive an efficiency bound, they give a clear intuition of the reason why their block-based estimator is nearly efficient. Within each block, one computes the maximum likelihood estimator of the variance of returns seen as approximately homoskedastic within the blocks. Then the sum across blocks of power p of these estimators delivers a smaller asymptotic variance than the naive estimator $\hat{\psi}_{g_p}$ *when both block size and number of blocks go to infinity. In this sense, their estimator is* nearly *efficient in the sense of Section 2.4 below. Recently, Jacod and Rosenbaum (2013) introduced an estimator that also has a limiting variance (2.3.10). Indeed, consider their Theorem 3.2 with the notation* $d = 1$, $t = 1$, $s = u$, $g(c) = c^p$, $s_s = \sigma^2(u)$. *Then, their limiting variance (3.12) leads to*

$$\int_{u=0}^{1} p\sigma_0^{2(p-1)}(u)p\sigma_0^{2(p-1)}(u)2\sigma^4(u)\,du = 2p^2\int_{u=0}^{1}\sigma_0^{4p}(u)\,du,$$

which equals (2.3.10) for the regularly-spaced data case, i.e., $T' = 1$. *Jacod and Rosenbaum (2013) note that their estimator is efficient at, what they call, the constant volatility toy model, i.e.,* $\sigma(u) = \sigma$. *We actually show that their estimator achieves the efficiency bound also at non-constant volatility within our nonparamet-*

[8]The link between their and our notation is $T = 1$, $r = 2p$, and $H(u) = \int_{v=0}^{u} T'(v)^{-1}\mathrm{d}v$.

ric model.

Example 2. Laplace transform: *Consider, for some given $s \in \mathbb{R}_+$, the transformation $g_s(\sigma^2) = \exp(-s\sigma^2)$ which leads to the efficiency bound*

$$V(g_s) = 2s^2 \int_0^1 \exp\left(-2s\sigma_0^2(u)\right) \sigma_0^4(u) du. \tag{2.3.15}$$

Todorov and Tauchen (2012) estimate $\psi_{g_s}(\sigma^2)$ with $\omega(u) \equiv 1$) by the realized Laplace transform

$$\hat{\psi}_{g_s} = \sum_{i=1}^{n} (t_{i,n} - t_{i-1,n}) \cos\left(\sqrt{2s}\frac{R_{i,n}}{\sqrt{t_{i,n} - t_{i-1,n}}}\right), \tag{2.3.16}$$

which gives them (see their Theorem 1 with $u = v$), the asymptotic variance

$$V = 2\int_0^1 \exp\left(-2s\sigma^2(u)\right) \left[\frac{\exp(s\sigma^2(u)) - \exp(-s\sigma^2(u))}{2}\right]^2 du. \tag{2.3.17}$$

Using the series expansion

$$\frac{\exp(x) - \exp(-x)}{2} = x\left[1 + \sum_{j=1}^{\infty} \frac{x^{2j}}{(2j+1)!}\right]$$

we find

$$V = 2s^2 \int_0^1 \exp\left(-2s\sigma^2(u)\right) \sigma_0^4(u) \left[1 + \sum_{j=1}^{\infty} \frac{\left(s\sigma^2(u)\right)^{2j}}{(2j+1)!}\right]^2 du.$$

As all terms in the series are nonnegative, the realized Laplace transform does not attain the efficiency bound for estimation of the Laplace transform $\psi_{g_s}(\sigma^2)$ (unless $s = 0$), although in applications the difference may be small. In Section 2.4, we provide a nearly efficient estimator which is also applicable in this case.

Even though they only address the efficiency issue "in the toy model" of i.i.d. homoskedastic normal returns (constant volatility σ), Jacod and Rosenbaum (2013) give a general statement about inefficiency of naive sample counterparts like realized power variation or the realized Laplace transform. They explain that efficiency requires in general to use "estimators for the spot volatility and approximating the integral $\psi_g(\sigma^2)$ by Rieman sums, in which the spot volatility is replaced by its estimator". As they rightly mention, this idea can be seen as a generalization of the block-based estimation idea in Mykland and Zhang (2009). It should be added that in earlier work, Kristensen (2010) had a germane idea by plugging in a kernel-based estimator of spot volatility. Both Jacod and Rosenbaum (2013), with a block-based approach, and Kristensen (2010), with a kernel-based approach, get asymptotic variances that attain the efficiency bound, even though they don't present them as the efficiency bound. Also, these papers do not consider the effect of irregular sampling. We turn to this issue now.

2.3.2 Irregular sampling without weighting

With a flat weighting function but possibly irregular sampling, we get the efficiency bound

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\right)^2 \frac{\sigma_0^4(u)}{T'(u)}\mathrm{d}u \tag{2.3.18}$$

that is new in the literature. In the particular case of integrated volatility ($g_1(\sigma^2) = \sigma^2$), the efficiency bound above corresponds to the asymptotic variance of realized variance, first derived by Barndorff-Nielsen and Shephard (2002a), at least when $H'(u) = 1/T'(u)$, a.e. This is a confirmation of the aforementioned intuition about the linearity of the transformation g_1, that allows to commute its integration with the smoothing operation. This efficiency result must be contrasted with two seemingly opposite claims in the literature.

First, Hayashi et al. (2011) claim (see their p. 1206) that "the realized volatility is asymptotically efficient only when the sampling scheme is asymptotically a regular sampling", that is $T'(u) \equiv 1$. However, our result proves the semiparametric efficiency of realized variance, even with irregular sampling. This seemingly contradictory result comes from the fact that Hayashi et al. (2011) define efficiency through the toy parametric model $\sigma^2(u) \equiv \sigma^2$, constant. Their remark 1thus means that the model with irregular sampling is not adaptive; see Section 2.5 for a more comprehensive discussion.

Second, as in the case of regular sampling, the case $p = 1$ is obviously the only one for which ψ_{g_p} is efficiently estimated by its naive sample counterpart. While Mykland and Zhang (2009) study the estimation of power variation only in the case of regular sampling, they refer to Mykland and Zhang (2012) for an extension to irregular sampling. Mykland and Zhang (2012) do discuss the interaction between the block-based estimation strategy and the irregular sampling scheme, but they don't provide explicitly the semiparametric efficiency bound or the way to reach it.

2.4 A nearly efficient estimator

The advantage of a LAN result as in Theorem 2.2 is that it indicates ways to construct efficient estimators. More precisely, the likelihood expansion, which in our case is based on the RV^* process, provides an asymptotically sufficient statistic for the parameter of interest, i.e., the path of σ^2. Hence, it is natural to base estimators on RV^*, a route we will follow in this section. This will lead to, what we call, a *nearly* efficient estimator. That is, we will base our estimator on a fixed smoothing kernel K (to be introduced formally below). The estimator then is nearly efficient in the sense that its limiting variance can get arbitrarily close to the nonparametric lower bound (2.3.10), by taking a kernel K close to the point mass at zero. In this sense, the Mykland and Zhang (2009) estimator is also nearly efficient, while Jacod and Rosenbaum (2013) provide explicit convergence rates for their smoothing parameter to achieve simultaneous convergence, i.e., (full) efficiency. However, note that in both cases (near) efficiency is reached only with asymptotically regular sampling schemes — the case where the two sequences RV_n and RV_n^* are asymptotically equivalent. Our near efficiency result below will be more general since it applies to irregular, even random (albeit predictable), observation times. The trick is to base estimation on a smoothed version of the process RV_n^*.

To get the main intuition, it's worth starting with the kernel-based estimator put forward by Kristensen (2010). That paper introduces a kernel-based estimator of spot volatility

$$\begin{aligned}\hat{\sigma}^2(u) &= \sum_{i=1}^{n} k_h(t_{i,n}-u)R_{i,n}^2 = \int_0^1 k_h(v-u)\mathrm{d}RV_n(v) \\ k_h(v) &= \frac{1}{h}k(v/h), \qquad \int_{-\infty}^{+\infty} k(v)\mathrm{d}v = 1.\end{aligned} \tag{2.4.19}$$

Kristensen (2010) actually uses standard realized variation RV, instead of RV^*, but, as mentioned before, the two are asymptotically equivalent in case of regular sampling. Kristensen (2010) shows that, under some regularity conditions (including continuous differentiability of the kernel function k)

$$\sup_{a\leq u\leq 1-a} \left|\hat{\sigma}^2(u)-\sigma^2(u)\right| = O_P(h^m) + O_P(\log(n)/\sqrt{nh}), \tag{2.4.20}$$

as $h \downarrow 0$, $a \downarrow 0$, and $a/h \to 0$, where the spot volatility function σ is assumed to be m times differentiable. We use Kristensen (2010)'s intuition and define a smoothed version of realized variance, which is pathwise differentiable, as

$$RV_n^S(u) = \int_{v=0}^{1} K(u-v)\mathrm{d}RV_n^*(v), \quad 0 \leq u \leq 1, \tag{2.4.21}$$

where K satisfies the following condition.

Assumption 2.8. *The kernel K is a non-negative real-valued function on* $[-1, 1]$ *which is twice continuously differentiable with K' and K'' bounded.*

One can take, for instance, K as the cumulative distribution function of a probability distribution whose density correponds to the function k defined in (2.4.19) and assume it to be continuously differentiable. In particular, we will keep the intuition that, with regular sampling and a probability distribution converging towards pointmass at zero (irrespective of its definition through the density function k or through the cumulative distribution function K), we would have $\hat{\sigma}^2(u)$ close to $\sigma^2(u)$ and, for the same reason, $RV_n^S(u)$ close to $RV_n^*(u)$. It is then natural to extend also Kristensen (2010)'s idea of a plug-in estimator for $\psi_g(\sigma^2)$, defined (up to a local bias correction) as

$$\hat{\psi}_g = \int_0^1 g(u, \hat{\sigma}^2(u))\mathrm{d}u. \tag{2.4.22}$$

In our notation, this would give in the case of regular sampling

$$\hat{\psi}_g(K|\text{reg}) = \int_0^1 g(u, RV_n^{S\prime}(u))du \tag{2.4.23}$$

However, this estimator would not be efficient in general with irregular sampling. To see that, it is worth recalling that, as stressed by Hayashi et al. (2011), see their p. 1205, the maximum likelihood estimator of σ^2, in the toy parametric model $\sigma^2(u) \equiv \sigma^2$ constant, is

$$\frac{1}{N_n} \sum_{i=1}^{n} \frac{R_{i,n}^2}{t_{i,n} - t_{i-1,n}}. \tag{2.4.24}$$

This suggests that in the plug in estimator (2.4.22), the instantaneous variance should be divided by the corresponding time increment, that is multiplied by $T_n'(u)$. This is precisely what RV_n^*, or rather its smoothed version $RV_n^S(u)$, does. Hence, we propose the following estimator.

$$\hat{\psi}_g(K) := \int_{u=0}^1 g\left(u, \frac{RV_n^{S\prime}(u)}{T_n'(u)}\right) \mathrm{d}u. \tag{2.4.25}$$

As already explained, we are not able, under our weak assumptions, to do an analysis as Kristensen (2010). He proves, still in the case of regular sampling, that his estimator (2.4.22), with a convenient bandwidth sequence h_n converging to zero and a convenient bias correction, is consistent and asymptotically normal with a variance that coincides with our efficiency bound $V(g)$. By contrast, our approach amounts to first considering the limiting behavior of the estimator (2.4.26 for fixed kernel K (Theorem 2.4 below). Then, in a second stage (Proposition 2.1 below) we derive the asymptotic behavior of the bias and variance of our estimator when the kernel itself converges weakly to the point mass at zero, that is

$$K(u) \to K^0(u) \text{ for all } u \neq 0 \text{ with } K^0(u) = 1_{\{[0,\infty)\}}(u).$$

It is important to stress that the validity of these two results is (much) more general than the set-up presented in this paper to derive the efficiency bound: It's only Condition 1 that is needed, not the sufficient conditions as provided in Lemma 2.1.

Theorem 2.4. *Under Condition 1 and Assumption 2.7-2.8, we have, under* $\mathbb{P}^{(n)}_{\sigma_0^2}$,

$$\begin{aligned} &\sqrt{n}\left(\hat{\psi}_g(K) - \psi_g\left(\sigma_0^2|K, T_n\right)\right) \\ &\quad \to_L \quad \sqrt{2}\int_{u=0}^{1} \frac{\partial g}{\partial \sigma^2}\left(u, T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)dT(w)\right) \\ &\qquad\qquad \times T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}dW(w)du, \end{aligned} \tag{2.4.26}$$

where

$$\psi_g\left(\sigma^2|K,T\right) = \int_{u=0}^{1} g\left(u, -T'(u)^{-1}\int_{w=0}^{1} T'(w)\sigma^2(w)dK(u-w)\right)du. \tag{2.4.27}$$

Theorem 2.4 gives the limiting behavior for fixed kernel K. However, to get a consistent estimator of the true unknown value of $\psi_g\left(\sigma_0^2\right)$, we need to consider estimators computed with a kernel K close to K^0. Thus, we consider, subsequently, sequences K_n such that $\lim_{n\to\infty} K_n(u) = K^0(u)$, for all $u \neq 0$. By abuse of language, we will say that the sequence K_n converges weakly to the pointmass at zero. Our near efficiency result is then formalized in the following proposition.

Proposition 2.1. *The limit $\psi_g\left(\sigma^2|K,T\right)$ defined in (2.4.27) satisfies*

$$\psi_g\left(\sigma_0^2|K,T\right) \to \psi_g\left(\sigma_0^2\right), \tag{2.4.28}$$

as K converges weakly to the point mass at zero. Moreover, the limiting variance of (2.4.26) equals

$$\begin{aligned}
&2\int_{w=0}^{1}\left[\int_{u=0}^{1} T'(u)^{-1}\frac{\partial g}{\partial\sigma^2}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)dT(v)\right)dK(u-w)\right]\sigma_0^4(w)T'(w)dw \\
&\to\ 2\int_{w=0}^{1}\left(\frac{\partial g}{\partial\sigma^2}\left(w,\sigma_0^2(w)\right)\right)^2\frac{\sigma_0^4(w)}{T'(w)}dw,
\end{aligned} \tag{2.4.29}$$

as K converges weakly to the point mass at zero.

In view of the above proposition, we call our estimator *nearly* efficient: the limiting distribution for given kernel K can get arbitrarily close to the lower bound. This is, clearly, a weaker result than the estimator provided in Jacod and Rosenbaum (2013). However, we obtain it within a framework allowing for random, though predictable, irregularly spaced observation times. This complicates the analysis significantly as also can be seen from a closer inspection of Theorem 2.4 where the centering in the central limit theorem is at $\psi_g\left(\sigma_0^2|K,T_n\right)$, i.e., using the observation times represented by T_n. It will, in general, not be the case that $\psi_g\left(\sigma_0^2|K,T\right)$ and $\psi_g\left(\sigma_0^2|K,T_n\right)$ differ in the order of $o_P(n^{-1/2})$ only, unless K tends to the pointmass at zero at an appropriate rate. We leave such a construction for further research.

2.5 On parametric information about the volatility process

The lower bound in Theorem 2.3 constitutes the *nonparametric* lower bound for estimating integrated functions of variance. In this section we focus on the example of power variation, namely $g_p(\sigma^2) = \sigma^{2p}$ and the simplified notation $\psi_{g_p} = \int_0^1 \sigma_0^{2p}(u)\mathrm{d}u$. Instead of considering a nonparametric lower bound, practitioners may prefer to specify a parametric functional for $\sigma^2(u)$. In this section, we study the effect of imposing such parametric information about the time-variation of σ^2, that is, when $\sigma^2(u) = \sigma^2(u|\theta)$ for some (sufficiently smooth) parametrization $\theta \mapsto \sigma^2(\cdot|\theta)$. We will show, at odds with what is sometimes considered common wisdom in this setting, that there are gains from such added information and these efficiency gains can be large.

The case of assuming constant volatility, that is $\sigma^2(u|\theta) = \theta > 0$ for $u \in [0, 1]$, has been studied extensively in Xiu (2010) and Jacod and Rosenbaum (2013). In the case of equally spaced data, and still ignoring market microstructure noise, the Gaussian QMLE estimator for θ equals $RV_n(1)$. As a result, its limiting variance equals $2\int_0^1 \sigma_0^4(u)\mathrm{d}u = 2\theta_0^2$ and thus the parametric and nonparametric lower bounds coincide.

The results above, in particular that the parametric and nonparametric bounds for estimating integrated variance are equal for the simplest case of constant volatility, should not be interpreted as that no inference gains are possible if parametric information about the form of the volatility process is available. In case such information is available, the MLE estimator, and thus the QMLE estimator, may have a variance strictly smaller than $2\int_0^1 \sigma_0^4(u)\mathrm{d}u$ at data generating processes for which the true underlying volatility σ_0^2 is not constant.

Consider a general parametric model $\sigma^2(\cdot|\theta)$ which is assumed to be sufficiently smooth such that the induced maximum-likelihood estimator satisfies the standard asymptotic expansion

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}^{(n)} - \theta_0\right) &= \frac{-1}{2\sqrt{n}} \sum_{i=1}^{n} \left[\frac{R_{i,n}^2}{\left(t_{i,n} - t_{i-1,n}\right)^2 / \int_{t_{i-1,n}}^{t_{i,n}} \sigma^{-2}(u|\theta_0)\mathrm{d}u} - 1 \right] \\ &\quad \times I(\theta_0)^{-1} \frac{\partial}{\partial\theta} \log \sigma^2\left(t_{i-1,n}|\theta_0\right) + o_{\mathbb{P}}(1),\end{aligned} \tag{2.5.1}$$

under $\mathbb{P}^{(n)}_{\sigma^2(\cdot|\theta_0)}$, with Fisher Information

$$I(\theta) = \frac{1}{2}\int_{u=0}^{1}\left(\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\right)\left(\frac{\partial}{\partial\theta'}\log\sigma^2(u|\theta)\right)\mathrm{d}T(u). \tag{2.5.2}$$

Consequently, the implied estimator for $\psi_p(\theta) = \int_0^1 \sigma^{2p}(u|\theta)\mathrm{d}u$ has limiting variance $\dot{\psi}_p(\theta_0)' I(\theta_0)^{-1}\dot{\psi}_p(\theta_0)$, under $\mathbb{P}^{(n)}_{\sigma^2(\cdot|\theta_0)}$, with $\dot{\psi}_p(\theta) = p\int_0^1 \sigma^{2p}(u|\theta)\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\mathrm{d}u$. We now have the following result.

Theorem 2.5. *Let Θ be an open subset of $\mathbb{R}^k$. Consider a parametric model $\{\sigma^2(\cdot|\theta) : \theta \in \Theta\}$ for which $\psi_p(\theta)$ is differentiable and the maximum-likelihood estimator satisfies (2.5.1)-(2.5.2). Then we have the following.*

(i) The limiting variance of the maximum likelihood estimator is at most (2.3.10).

(ii) Equality holds in case $\sigma^{2p}(\cdot|\theta)/T'(u)$ is piecewise constant with at most $k = \dim(\theta)$ different values, i.e., in case we can partition $[0,1]$ into subsets $A_1, \ldots, A_k$ such that

$$\frac{\sigma^{2p}(u|\theta)}{T'(u)} = \sum_{j=1}^{k}\exp(d(\theta))_j I_{\{u\in A_j\}}, \tag{2.5.3}$$

for some C_1-diffeomorphism $d : \Theta \to \mathbb{R}^k$.

Proof. In order to prove (i), fix $\theta \in \Theta$ and project $\sigma^{2p}(u|\theta)/T'(u)$ on the space spanned by the elements of $\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)$, i.e., write $\sigma^{2p}(u|\theta)/T'(u) = \beta'\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta) + \eta(u)$ where $\int_{u=0}^1 \eta(u)\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\mathrm{d}T(u) = 0$. Plugging this decomposition in the limiting variance of the maximum likelihood estimator yields

$$
\begin{aligned}
\dot{\psi}_p(\theta_0)' I(\theta_0)^{-1} \dot{\psi}_p(\theta_0) &= 2p^2 \int_0^1 \frac{\sigma^{2p}(u|\theta)}{T'(u)} \frac{\partial}{\partial \theta'} \log \sigma^2(u|\theta) \mathrm{d}T(u) \\
& \quad \times \left[\int_{u=0}^1 \left(\frac{\partial}{\partial \theta} \log \sigma^2(u|\theta) \right) \left(\frac{\partial}{\partial \theta'} \log \sigma^2(u|\theta) \right) \mathrm{d}T(u) \right]^{-1} \\
& \quad \times \int_0^1 \frac{\sigma^{2p}(u|\theta)}{T'(u)} \frac{\partial}{\partial \theta} \log \sigma^2(u|\theta) \mathrm{d}T(u) \\
&= 4p^2 \beta' I(\theta) \beta \leq 4p^2 \beta' I(\theta) \beta + 2p^2 \int_{u=0}^1 \eta^2(u) \mathrm{d}T(u) \\
&= 2p^2 \int_{u=0}^1 \left(\frac{\sigma^{2p}(u)}{T'(u)} \right)^2 \mathrm{d}T(u) = 2p^2 \int_{u=0}^1 \frac{\sigma^{4p}(u)}{T'(u)} \mathrm{d}u.
\end{aligned}
$$

Concerning Part (ii), note that equality holds if and only if $\eta = 0$, that is, if and only if $\sigma^{2p}(u|\theta)/T'(u) = \beta' \frac{\partial}{\partial \theta} \log \sigma^2(u|\theta) = (p^{-1}\beta)' \frac{\partial}{\partial \theta} \log(\sigma^{2p}(u|\theta)/T'(u))$ for some $\beta \in \mathbb{R}^k$ (possibly dependent on θ, but not on u). Clearly, (2.5.3) indeed provides a sufficient condition for equality of both bounds by taking $p^{-1}\beta = \left[\frac{\partial}{\partial \theta'} d(\theta)\right]^{-1} \exp(d(\theta))$, where the exponential is applied componentwise. ■

The above theorem has an interesting implication, even in case of equally-spaced data, i.e., $T'(u) = 1$ for all $u \in [0,1]$. Then, the above theorem implies that for piecewise constant parametrizations of the volatility function σ (still with at most $k = \dim(\theta)$ different values) the parametric and nonparametric lower bounds coincide. In that case, for $p = 1$, we thus have that realized variance is even a parametrically efficient estimator of integrated variance.

However, for other (not piecewise constant) parametric specifications, this is not true. For example, consider the case where a researcher would specify $\sigma^2(u|\theta) = \exp(\theta_1 + \theta_2 u)$, $u \in [0,1]$, still with equally spaced data. The Figure 2.1 shows the ratio of the nonparametric and the parametric lower bounds for estimating integrated power variance for $p = 1$, 2, and 3 in the simple case where $\theta_1 = 0$ and $-5 \leq \theta_2 \leq 5$. Focusing on estimating integrated variance, the inference gain from the information on the parametric form of the volatility function exceeds 15% when

$\theta_2 = 5$. Moreover, the gain from information goes up considerably as p increases. For example, the ratio of the nonparametric and the parametric lower bounds is as much as 1.71 for estimating integrated quarticity ($p = 2$) and 2.30 for estimating integrated third power variance ($p = 3$) when $\theta_2 = 5$. These values are both theoretically and practically significant.

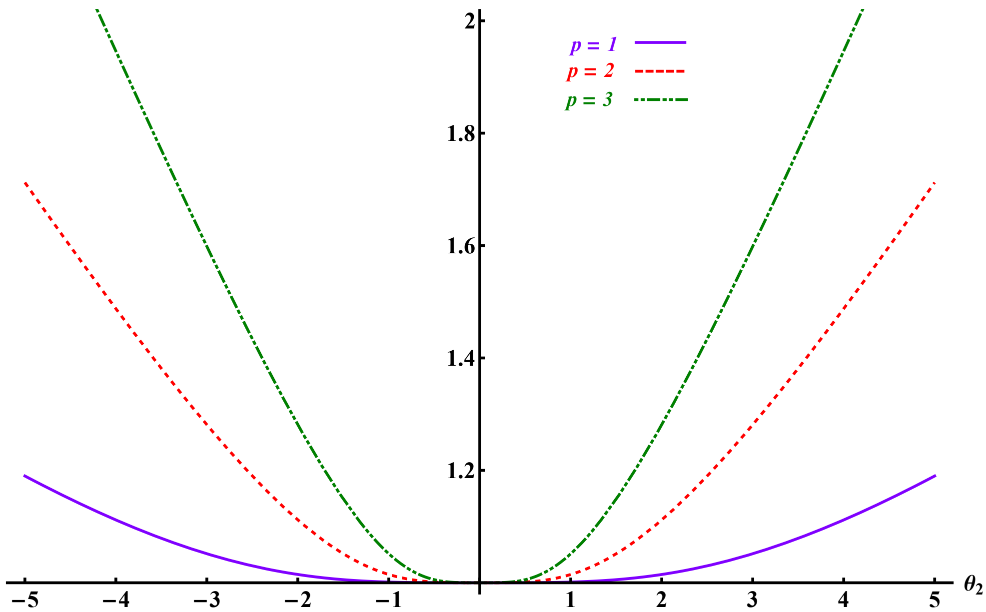

Figure 2.1: The ratio of nonparametric and parametric lower bounds for estimating integrated powers of variance where $\sigma^2(u|\theta) = \exp(\theta_1 + \theta_2 u)$, $u \in [0, 1]$, with equally spaced data, i.e., $T'(u) = 1$. This figure considers the powers $p = 1, 2$ and 3 with $\theta_1 = 0$ and $-5 \leq \theta_2 \leq 5$.

Clearly, the appropriateness of specifying a parametric model for the time-evolution of intraday volatility is generally an empirical question, with the classical trade-off between possible misspecification and efficiency.

2.6 Conclusions

The results in the present paper complement those of Reiß (2011) by focusing on nonparametric lower bounds for integrated powers of volatility, in the absence of market micro structure noise. In line with Clément et al. (2013), we find locally and asymptotically normal limiting experiments at rate $\sqrt{n}$, i.e., the limiting experiments with and without microstructure noise are materially different. Unlike Clément et al. (2013), we focus on the pathwise properties of the volatility process that are needed to obtain this limit. Using these results, we establish the (near) efficiency of the estimator put forward in Mykland and Zhang (2009). Moreover, we demonstrate the efficiency of the Jacod and Rosenbaum (2013) estimator, also at non-constant volatility within the nonparametric model.

Second, we detail the role of random, though predictable, unequally spaced observation times. We establish precisely how these affect the efficiency bounds. For integrated variance, classical realized variance is efficient, also under these irregular sampling schemes. For higher powers, we provide an estimator that is nearly efficient, i.e., whose limiting variance can get arbitrarily close to the nonparametric lower bound.

Finally, we provide a simple condition under which there are no gains from assuming a parametric specification of the volatility function. This is important as, in applied work, one may prefer the risk of misspecification of a parametric form over the loss of efficiency of fully nonparametric procedures. We show that, with the exception of some very particular volatility functions in relation to the observation scheme, significant efficiency gains are possible in general.

Appendix

2.A Some lemmas and details on Locally Bounded Variance

We start by a lemma that bounds the variance of the inverse of positive random variables.

Lemma 2.6. *Let $X \geq c > 0$, then $V\{X^{-1}\} \leq c^{-4} Var\{X\}$.*

Proof. We have the following inequalities:

$$\begin{aligned} \mathrm{V}\left\{\frac{1}{X}\right\} &\leq \mathrm{E}\left(\frac{1}{X} - \frac{1}{\mathrm{E}\{X\}}\right)^2 \\ &\leq c^{-4}\mathrm{E}\left(X - \mathrm{E}\{X\}\right)^2, \end{aligned}$$

as, for $x, y > c$, $|x^{-1} - y^{-1}| \leq |x - y|/c^2$ since the derivative of $x \mapsto x^{-1}$ is bounded by c^{-2}. ■

The bound (2.2.4) relates the assumption of locally bounded variance for the sample paths of a stochastic process to its quadratic variation. We know that if $\{X(t) : 0 \leq t \leq 1\}$ is a semi-martingale, then its quadratic variation is defined as the probability limit

$$P \lim_{n\to\infty} \sum_{i=1}^{N_n} \left[X(t_{i,n}) - X(t_{i-1,n})\right]^2 = \langle X, X \rangle_1 , \tag{2.A.1}$$

see, e.g., Protter (1995), Theorem II.22. It is a common assumption to consider the volatility process to be a semimartingale or, even smoother, a Fractional Brownian Motion. Thus, at least for convenient sampling schemes with

$$\left[\max_{t_{i-1,n}\leq u\leq t_{i,n}} X(u) - \min_{t_{i-1,n}\leq u\leq t_{i,n}} X(u)\right]^2 = \left[X(t_{i,n}) - X(t_{i-1,n})\right]^2,$$

the existence of quadratic variation may ensure the convergence in probability of an upper bound of $\sum_{i=1}^{N_n} \mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right)$. However, convergence in probability does not ensure the required almost sure boundedness of $\sum_{i=1}^{N_n} \mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right)$. For instance, it is known that if $X_{t\in[0,1]}$ is a Brownian motion, $\sup\sum_{i=1}^{N_n}\left[X(t_{i,n}) - X(t_{i-1,n})\right]^2$, where the supremum is computed over all possible conformable sampling schemes, is almost surely infinite (see, e.g., the remark below Definition I(2.3) in Revuz and Yor (1991)). One way to circumvent the above issue would be to ensure that the convergence in (2.A.1) is not only in probability but also almost surely. Almost sure convergence would hold if $\{X(t) : 0 \leq t \leq 1\}$ is a Brownian motion and we consider sequences of sampling schemes that are refining in the sense that

$$\{t_{i,n} : \ i = 1, \ldots, N_n\} \subset \{t_{i,n+1} : \ i = 1, \ldots, N_{n+1}\}, \tag{2.A.2}$$

see Protter (1995), Theorem I.28. Nevertheless, fortunately, for a process X that is an $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-Brownian Motion in the sense of Definition III(2.20) in Revuz and Yor (1991), we can show directly that it is of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance by applying a convenient law of large numbers to an upper bound of the process $\mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right)$.

Proposition 2.2. *If the process* $X = \{X(t) : 0 \leq t \leq 1\}$ *is a* $\{\mathcal{F}_t\}_{0\leq t\leq 1}$*-Brownian Motion, its sample paths are almost surely of locally* $\{\mathcal{F}_t\}_{0\leq t\leq 1}$*-bounded variance.*

Proof. First, note that we can assume without loss of generality that $N_n = n$. Indeed, if $N_n < n$, we can always complete the sampling scheme as follows

$$i \geq N_n \Rightarrow t_{i+1,n} = t_{i,n}, \tag{2.A.3}$$

so that, for any f,

$$\sum_{i=1}^{N_n} \mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) = \sum_{i=1}^{n} \mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right). \tag{2.A.4}$$

Moreover, the sampling scheme extended by 2.A.3 still fulfills Assumptions 2.1 and 2.2.

We use the following lemma.

Lemma 2.7. *If* $X = \{X(t) : 0 \le t \le 1\}$ *is a* $\{\mathcal{F}_t\}_{0\le t\le 1}$*-Brownian Motion and the sampling scheme* $(t_{i,n})_{0\le i\le n}$ *satisfies Assumption 2.2, then we can write*

$$V\left(X|_{t_{i-1,n}}^{t_{i,n}}\right) = (t_{i,n} - t_{i-1,n})\, V\left(Z_n^{(i)}\big|_0^1\right), \tag{2.A.5}$$

where $Z^{(i)}$, $i = 1, \ldots, n$, *are independent standard Brownian motions on* $[0, 1]$, *with with* $Z_n^{(i)}$ *independent of* $\mathcal{F}_{t_{i-1,n}}$.

Proof. The only non-trivial case is when $\Delta t_{i,n} = t_{i,n} - t_{i-1,n} > 0$. In that case, define $Z_n^{(i)}$ by

$$Z_n^{(i)}(s) = \frac{X(t_{i-1,n} + s\Delta t_{i,n}) - X(t_{i-1,n})}{\sqrt{\Delta t_{i,n}}}, \quad 0 \le s \le 1.$$

Now, note that since $X = \{X(t) : 0 \le t \le 1\}$ is a $\{\mathcal{F}_t\}_{0\le t\le 1}$-Brownian Motion and the sampling scheme $(t_{i,n})_{0\le i\le n; n\ge 1}$ is $\{\mathcal{F}_t\}_{0\le t\le 1}$-measurable, $Z_n^{(i)}, i = 1, \ldots, n$, are independent stochastic processes on $[0, 1]$. The fact that the $Z_n^{(i)}$'s are standard Brownian motions follows from the self-similarity property of Wiener processes as well as from the fact that $\Delta t_{i,n}$ is known at time $t_{i-1,n}$.

Now, observe that since $\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)$ equals $\mathrm{V}(f(U))$ with U uniformly distributed over $[t_{i-1,n}, t_{i,n}]$, we find

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) = \int_0^1 f(t_{i-1,n} + s\Delta t_{i,n})^2 \mathrm{d}s - \left[\int_0^1 f(t_{i-1,n} + s\Delta t_{i,n})\mathrm{d}s\right]^2.$$

Then,

$$\begin{aligned}
\mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right) &= \int_0^1 \left[X_{t_{i-1,n}} + \sqrt{\Delta t_{i,n}} Z_n^{(i)}(s)\right]^2 \mathrm{d}s - \left[\int_0^1 X_{t_{i-1,n}} + \sqrt{\Delta t_{i,n}} Z_n^{(i)}(s)\mathrm{d}s\right]^2 \\
&= X_{t_{i-1,n}}^2 + \Delta t_{i,n} \int_0^1 Z_n^{(i)}(s)^2 \mathrm{d}s + 2X_{t_{i-1,n}} \sqrt{\Delta t_{i,n}} \int_0^1 Z_n^{(i)}(s)\mathrm{d}s \\
&\quad - X_{t_{i-1,n}}^2 - \Delta t_{i,n} \left[\int_0^1 Z_n^{(i)}(s)\mathrm{d}s\right]^2 - 2X_{t_{i-1,n}} \sqrt{\Delta t_{i,n}} \int_0^1 Z_n^{(i)}(s)\mathrm{d}s \\
&= \Delta t_{i,n} \left\{\int_0^1 Z_n^{(i)}(s)^2 \mathrm{d}s - \left[\int_0^1 Z_n^{(i)}(s)\mathrm{d}s\right]^2\right\} \\
&= \Delta t_{i,n} \mathrm{V}\left(Z_n^{(i)}|_0^1\right).
\end{aligned}$$

■

We continue the proof of Proposition 2.2. It is sufficient to prove that

$$\sum_{i=1}^n \Delta t_{i,n} \mathrm{V}\left(Z_n^{(i)}|_0^1\right)$$

is almost surely bounded. From Cauchy-Schwarz, this follows as both

$$n\sum_{i=1}^n (\Delta t_{i,n})^2 \text{ and } n^{-1}\sum_{i=1}^n \mathrm{V}\left(Z_n^{(i)}|_0^1\right)^2$$

are bounded almost surely by Assumption 2.1 and a standard strong law of large numbers. ■

Reiß (2011) imposes smoothness conditions on the sample paths of volatility in terms of Hölder balls. Note that if for some $\alpha > 0$, $\sup_{u \neq v} |f(u) - f(v)| \,/\, |u - v|^\alpha \leq R$, we have from (2.2.3)

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) \leq \frac{1}{4} \sup_{t_{i-1,n} \leq u \neq v \leq t_{i,n}} [f(u) - f(v)]^2 \leq \frac{1}{4} R^2 (\Delta t_{i,n})^{2\alpha}.$$

As a result, in view of Assumption 2.1, f is of locally bounded variance for $\alpha \geq 1/2$ (apply, e.g., the Cauchy-Schwarz inequality and use $\sum(\Delta t_{i,n})^2 = O_P(n^{-1})$). Reiß

(2011) needs $\alpha \geq (1+\sqrt{5})/4 \approx 0.81$ in his main Theorem 6.2. This, again, shows that the analysis with and without (microstructure) noise, leads to materially different limiting experiments.

Given these considerations, we consider throughout that the assumption of a process being almost surely of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance is not overly restrictive. Whether, besides the cases discussed above, paths of general semimartingales are of locally bounded variance is still an open problem.

We also provide an auxiliary lemma to bound certain expressions that can be interpreted as expectations and arise in the various proofs.

Lemma 2.8. *Let X and Y be two strictly positive random variables bounded from above by M. Then, we have the following bounds*

$$
\begin{aligned}
0 \leq E\{X\} - \left(E\{X^{-1}\}\right)^{-1} &\leq M\sqrt{V\{X\}\,V\{X^{-1}\}}, && (2.A.6)\\
0 \leq E\{X^2\} - \left(E\{X^{-1}\}\right)^{-2} &\leq V\{X\} + 2M^2\sqrt{V\{X\}\,V\{X^{-1}\}}, && (2.A.7)\\
\left|(E\{XY\})^2 - (E\{X\})^2\,E\{Y^2\}\right| &\leq 4M^2\sqrt{V\{X\}\,V\{Y\}} + M^2\,V\{Y\} && (2.A.8)
\end{aligned}
$$

Proof. The left-hand side inequality of (2.A.6) is well-known. For the right-hand side observe

$$0 \geq 1 - \mathrm{E}\{X\}\,\mathrm{E}\{X^{-1}\} = \mathrm{Cov}\{X, X^{-1}\}$$

As the harmonic mean of X is bounded by M, (2.A.6) follows from Cauchy-Schwarz.

To prove (2.A.7), write

$$\mathrm{E}\{X^2\} - \left(\mathrm{E}\{X^{-1}\}\right)^{-2} = \mathrm{V}\{X\} + \left(\mathrm{E}\{X\} + \left(\mathrm{E}\{X^{-1}\}\right)^{-1}\right)\left(\mathrm{E}\{X\} - \left(\mathrm{E}\{X^{-1}\}\right)^{-1}\right).$$

Finally, concerning (2.A.8), observe

$$
\begin{aligned}
&\left|(\mathrm{E}\{XY\})^2 - (\mathrm{E}\{X\})^2\,\mathrm{E}\{Y^2\}\right| \\
&= \left|(\mathrm{Cov}\{X,Y\})^2 + 2\mathrm{E}\{X\}\,\mathrm{E}\{Y\}\,\mathrm{Cov}\{X,Y\} - (\mathrm{E}\{X\})^2\,\mathrm{V}\{Y\}\right| \\
&\leq |\mathrm{Cov}\{X,Y\}\,[\mathrm{Cov}\{X,Y\} + 2\mathrm{E}\{X\}\,\mathrm{E}\{Y\}]| + M^2\mathrm{V}\{Y\} \\
&\leq 2M^2\,(\mathrm{V}\{X\}\,\mathrm{V}\{Y\})^{1/2} + M^2\mathrm{V}\{Y\}.
\end{aligned}
$$

■

2.B Proofs

2.B.1 An asymptotically equivalent model

The specification (2.2.7) is classical in applications, but inconvenient for the likelihood calculations underlying the proof of the LAN property. We, therefore, introduce here an alternative volatility specification, using harmonic means instead of arithmetic means, and show that both define asymptotically the same statistical experiments.

Thus, consider the alternative specification

$$\frac{1}{\sigma^2(t_{i-1,n}, t_{i,n})} = \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} \sigma^{-2}(u)\mathrm{d}u. \tag{2.B.1}$$

In this case, we write

$$
\begin{aligned}
\sigma^2(t_{i-1,n}, t_{i,n}) &= H_\sigma(t_{i-1,n}, t_{i,n}) \\
&:= \left[\frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} \sigma^{-2}(u)\mathrm{d}u\right]^{-1}.
\end{aligned}
$$

This specification differs from the arithmetic mean specification considered in (2.2.7)

$$\sigma^2(t_{i-1,n}, t_{i,n}) = \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} \sigma^2(u)\mathrm{d}u = A_\sigma(t_{i-1,n}, t_{i,n}).$$

The same functional parameter σ^2 will, in general, not produce the same value for the arithmetic and harmonic means. It is well known that the arithmetic mean $A_\sigma(t_{i-1,n}, t_{i,n})$ and the harmonic mean $H_\sigma(t_{i-1,n}, t_{i,n})$ coincide only if the function σ^2 is constant over the interval $(t_{i-1,n}, t_{i,n}]$. However, as the length of each observation interval $(t_{i-1,n}, t_{i,n}]$ becomes asymptotically negligible, one may hope that a model defined in terms of the harmonic means $H_\sigma(t_{i-1,n}, t_{i,n})$ is equivalent to the financially more meaningful model defined in terms of the arithmetic means $A_\sigma(t_{i-1,n}, t_{i,n})$. Of course, this takes some smoothness on the volatility function σ. We formalize this in Proposition 2.3 below using precisely the smoothness condition of locally bounded variance for σ^2.

Recall that cadlag functions on a compact set like $[0, 1]$ are bounded, so that, under Assumption 2.5, both σ^2 and σ^{-2} will be bounded away from zero and infinity. Hence, for all $\sigma^2 \in \Xi$,

$$0 < m_\sigma := \inf_{0\le u\le 1} \left\{\sigma^2(u), \sigma^{-2}(u)\right\} \le M_\sigma := \sup_{0\le u\le 1} \left\{\sigma^2(u), \sigma^{-2}(u)\right\} < +\infty.$$

This allows us to uniformly control the difference between arithmetic and harmonic means of σ^2. The proof of the following result is immediate from Lemma 2.8.

Lemma 2.9. *For all $\sigma^2 \in \Xi$, we have*

$$0 \le A_\sigma(t_{i-1,n}, t_{i,n}) - H_\sigma(t_{i-1,n}, t_{i,n}) \le M_\sigma \sqrt{V\left(\sigma^2|_{t_{i-1,n}}^{t_{i,n}}\right)}\sqrt{V\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)}.$$

Lemma 2.9 allows us to study the asymptotic equivalence of the two following statistical experiments in the sense of Le Cam (1986).

Experiment 1. *Suppose that observed returns satisfy Assumption 2.3 with, for*

$i = 1, \ldots, N_n$,

$$\sigma^2(t_{i-1,n}, t_{i,n}) = A_\sigma(t_{i-1,n}, t_{i,n}), \tag{2.B.2}$$

for some $\sigma^2 \in \Xi$.

Experiment 2. *Suppose that observed returns satisfy Assumption 2.3 with, for* $i = 1, \ldots, N_n$,

$$\sigma^2(t_{i-1,n}, t_{i,n}) = H_\sigma(t_{i-1,n}, t_{i,n}), \tag{2.B.3}$$

for some $\sigma^2 \in \Xi$.

Observe that the Experiments 1 and 2 have the same parameter space Ξ. We establish asymptotic equivalence by showing below that Le Cam's deficiency pseudo-distance between the Gaussian experiments converges to zero (almost surely in the sampling times), when $n \to \infty$. Since we deal with Gaussian distributions, it is convenient to use the known fact (see, e.g., Nussbaum (1996), Formula (12)) that Le Cam's squared pseudo-distance is bounded by four times the value of the squared Hellinger distance between the corresponding densities. In order to show that this Hellinger distance converges to zero, we use Assumption 2.4 to conclude (see, e.g., Lemma 2.4. in Nussbaum (1996)) that the squared Hellinger distance between Experiments 1 and 2 is bounded by

$$2 \sum_{i=1}^{N_n} D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right),$$

where $P^A_{\sigma,i,n}$ and $P^H_{\sigma,i,n}$ are, respectively, the normal probability distribution with variance $A_\sigma(t_{i-1,n}, t_{i,n})\Delta t_{i,n}$ and that with variance $H_\sigma(t_{i-1,n}, t_{i,n})\Delta t_{i,n}$ and the means as in Assumption 2.3. Here, $D^2(P^A_{\sigma,i,n}; P^H_{\sigma,i,n})$ stands for the squared Hellinger distance between $P^A_{\sigma,i,n}$ and $P^H_{\sigma,i,n}$, conditionally on the observation times.

It turns out that the squared Hellinger distance between two normal distributions

$N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is easily computed (see, e.g., Brown and Low (1996)) as

$$D^2\left(N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)\right) = 2\left\{1 - \left[\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\right]^{1/2} \exp\left[-\frac{(\mu_1-\mu_2)^2}{4(\sigma_1^2+\sigma_2^2)}\right]\right\}. \quad (2.B.4)$$

We now have the following result.

Proposition 2.3. *Under Assumptions 2.1-2.5, the statistical Experiments 1 and 2 are asymptotically equivalent in the sense that, almost surely,*

$$\lim_{n\to\infty} \sum_{i=1}^{N_n} D^2(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}) = 0.$$

Proof. From (2.B.4) we have

$$D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right) = 2\left\{1 - \left[\frac{2\sigma_{i1}\sigma_{i2}}{\sigma_{i1}^2+\sigma_{i2}^2}\right]^{1/2} \exp\left[-\frac{(\mu_{i1}-\mu_{i2})^2}{4(\sigma_{i1}^2+\sigma_{i2}^2)}\right]\right\},$$

where

$$\begin{aligned} \sigma_{i1}^2 &= A_\sigma(t_{i-1,n}, t_{i,n}), \\ \sigma_{i2}^2 &= H_\sigma(t_{i-1,n}, t_{i,n}), \\ (\mu_{i1}-\mu_{i2})^2 &= \gamma^2(t_{i-1,n}, t_{i,n})(\sigma_{i1}^2 - \sigma_{i2}^2)^2 \Delta t_{i,n}. \end{aligned}$$

We rewrite

$$\frac{2\sigma_{i1}\sigma_{i2}}{\sigma_{i1}^2+\sigma_{i2}^2} = 1 - \frac{(\sigma_{i1}-\sigma_{i2})^2}{\sigma_{i1}^2+\sigma_{i2}^2} = 1 - \frac{(\sigma_{i1}^2-\sigma_{i2}^2)^2}{(\sigma_{i1}^2+\sigma_{i2}^2)(\sigma_{i1}+\sigma_{i2})^2},$$

and define

$$\varepsilon_{i,n} := \frac{(\sigma_{i1}^2-\sigma_{i2}^2)^2}{\sigma_{i1}^2+\sigma_{i2}^2},$$

so that

$$D^2\left(P^A_{\sigma,i,n};P^H_{\sigma,i,n}\right)=2\left\{1-\left[1-\frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{1/2}\exp\left[-\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\varepsilon_{i,n}\right]\right\}.$$

Observe that the derivative of the map $\varepsilon_{i,n}\mapsto D^2\left(P^A_{\sigma,i,n};P^H_{\sigma,i,n}\right)$, for given σ_{i1}, σ_{i2}, and γ, is

$$\begin{aligned}
&\left|\frac{\partial}{\partial\varepsilon_{i,n}}D^2\left(P^A_{\sigma,i,n};P^H_{\sigma,i,n}\right)\right|\\
&= \left|\left[1-\frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{-1/2}(\sigma_{i1}+\sigma_{i2})^{-2}+2\left[1-\frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{1/2}\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\right|\\
&\quad\times\exp\left[-\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\varepsilon_{i,n}\right]\\
&\le \left[1-\frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{-1}(\sigma_{i1}+\sigma_{i2})^{-2}+\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{2}\\
&= \frac{1}{(\sigma_{i1}+\sigma_{i2})^2-\varepsilon_{i,n}}+\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{2},
\end{aligned}$$

ans thus is bounded by, say, K as $(\sigma_{i1}+\sigma_{i2})^2-\varepsilon_{i,n}=2\sigma_{i1}\sigma_{i2}+4\sigma_{i1}^2\sigma_{i2}^2/(\sigma_{i1}^2+\sigma_{i2}^2)\ge 2m_\sigma$ and γ is bounded by Assmption 2.3.

Thus, by the mean-value theorem,

$$\begin{aligned}
\sum_{i=1}^{N_n}D^2\left(P^A_{\sigma,i,n};P^H_{\sigma,i,n}\right) &\le K\sum_{i=1}^{N_n}\varepsilon_{i,n}\\
&\le \frac{1}{2m_\sigma}\max_{1\le i\le N_n}\left|\sigma_{i1}^2-\sigma_{i2}^2\right|\sum_{i=1}^{N_n}\left|\sigma_{i1}^2-\sigma_{i2}^2\right|.
\end{aligned}$$

But we also know

$$\left|\sigma_{i1}^2-\sigma_{i2}^2\right|=|A_\sigma(t_{i-1,n},t_{i,n})-H_\sigma(t_{i-1,n},t_{i,n})|\le M_\sigma\sqrt{V\left(\sigma^2|_{t_{i-1,n}}^{t_{i,n}}\right)}\sqrt{V\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)},$$

so that by the assumption of locally bounded variance we obtain

$$\sum_{i=1}^{N_n} D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right) \leq K \sum_{i=1}^{N_n} \varepsilon_{i,n} \to 0, \text{ a.s.}$$

■

Proposition 2.3 establishes the asymptotic equivalence of using the arithmetic or harmonic average of spot variance in the description of our experiment.

2.B.2 Proof of Lemma 2.1.

First, we verify that the centering proposed in (2.3.3) and (2.3.4) is indeed appropriate. Observe, using (2.A.6),

$$\begin{aligned}
&\sup_{0\leq u\leq 1} \sqrt{n}\left|\sum_{t_{i,n}\leq u} \mathrm{E}_{\sigma_0^2}\left\{ R^2_{i,n}\middle| \mathcal{F}_{t_{i-1,n}}\right\} - \int_{v=0}^{u} \sigma_0^2(v)\mathrm{d}v\right| \\
&\leq \sqrt{n}\sum_{i=1}^{n}\left[\mu(t_{i-1,n},t_{i,n}) + \gamma(t_{i-1,n},t_{i,n})\frac{1}{(t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right]^2 (t_{i,n}-t_{i-1,n})^2 \\
&\quad + (t_{i,n}-t_{i-1,n})\left[(t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(v)\mathrm{d}v - \frac{1}{(t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right].
\end{aligned}$$

We show that both terms in the summation above converges to zero. The second term can be bounded, using (2.A.6) once more, as

$$\sqrt{n}\max_{i=1,\ldots,n}|t_{i,n}-t_{i-1,n}|\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\left(\mathrm{V}\left(\sigma_0^2\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\mathrm{V}\left(\sigma_0^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\right)^{1/2},$$

which converges to zero in view of Assumption 2.1, the Cauchy-Schwarz inequality, and Assumption 2.5. Concerning the first term, recall that μ, γ, and σ_0^2 are all bounded. Hence it suffices to study

$$\sqrt{n}\sum_{i=1}^{n}(t_{i,n}-t_{i-1,n})^2, \tag{2.B.5}$$

which also converges to zero in the view of Assumption 2.1.

With respect to RV^*, we note that without loss of generality we may actually study the slightly redefined version

$$RV_n^*(u) = \sum_{i=1}^{n} \frac{R_{i,n}^2}{n\left(t_{i,n} - t_{i-1,n}\right)} I\left\{t_{i,n} \leq u\right\}, \tag{2.B.6}$$

which differs at most $R_{i,n}^2/\left[n\left(t_{i,n} - t_{i-1,n}\right)\right] = O_{\mathbb{P}}(1/n)$ from (2.3.2). For this version we find, again using Lemma 2.A.6,

$$\begin{aligned}
&\sup_{0\leq u\leq 1} \sqrt{n}\left|\sum_{t_{i,n}\leq u} \frac{1}{n\left(t_{i,n} - t_{i-1,n}\right)} \mathrm{E}_{\sigma_0^2}\left\{\left. R_{i,n}^2\right| \mathcal{F}_{t_{i-1,n}}\right\} - \int_{v=0}^{u} \sigma_0^2(v)\mathrm{d}T_n(v)\right| \\
&\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mu(t_{i-1,n}, t_{i,n}) + \gamma(t_{i-1,n}, t_{i,n}) \frac{1}{\left(t_{i,n} - t_{i-1,n}\right)^{-1} \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v}\right]^2 \left(t_{i,n} - t_{i-1,n}\right) \\
&\quad + \left[\left(t_{i,n} - t_{i-1,n}\right)^{-1} \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^2(v)\mathrm{d}v - \frac{1}{\left(t_{i,n} - t_{i-1,n}\right)^{-1} \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v}\right].
\end{aligned}$$

Again, we show that both terms converge to zero. The second term can be bounded, in view of (2.A.6), by

$$\frac{1}{\sqrt{n}} \left\|\sigma_0^2\right\| \sum_{i=1}^{n} \left(\mathrm{V}\left(\sigma_0^2\Big|_{t_{i-1,n}}^{t_{i,n}}\right) \mathrm{V}\left(\sigma_0^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\right)^{1/2}, \tag{2.B.7}$$

which converges to zero in view of Assumption 2.5. For the first term, it is sufficient to consider

$$\sqrt{n} \max_{i=1,\ldots,n} \left|t_{i,n} - t_{i-1,n}\right|, \tag{2.B.8}$$

which converges to zero in view of Assumption 2.1.

Using the above results, we can now prove the claim by an application of Theorem VIII.3.33 in Jacod and Shiryaev (2002) to the exactly centered versions of the bivariate process (RV_n, RV_n^*). In the notation of Jacod and Shiryaev (2002), we have $k = i$, $t = u$ and $\sigma_t^n = nT_n(u)$ [recall $t_{i,n} = T_n^{-1}(i/n)$] and we consider the martingale difference sequence

$$
\begin{aligned}
U_i^n &= \sqrt{n}\begin{bmatrix} 1 \\ [n(t_{i,n}-t_{i-1,n})]^{-1}\end{bmatrix}\left(R_{i,n}^2 - \mathrm{E}_{\sigma_0^2}\left\{ R_{i,n}^2 \middle| \mathcal{F}_{t_{i-1,n}}\right\}\right) \qquad (2.B.9)\\
&= \sqrt{n}\begin{bmatrix} 1 \\ [n(t_{i,n}-t_{i-1,n})]^{-1}\end{bmatrix}\frac{(t_{i,n}-t_{i-1,n})^2}{\int_{v=t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\left(Z_{i,n}^2-1\right),
\end{aligned}
$$

with $Z_{i,n}$ i.i.d. standard normal. For the required Lindeberg condition, note that for $0 < M \leq \inf_{0\leq u\leq 1}\sigma_0^{-2}(u)$ we have

$$
|U_i^n| \leq \left\|\begin{bmatrix}\sqrt{n}\,(t_{i,n}-t_{i-1,n}) \\ 1/\sqrt{n}\end{bmatrix}\right\|\frac{\left|Z_{i,n}^2-1\right|}{M} = \sqrt{n\,(t_{i,n}-t_{i-1,n})^2+1/n}\,\frac{\left|Z_{i,n}^2-1\right|}{M}.
$$

Thus, for all $\varepsilon > 0$,

$$
\begin{aligned}
&\sum_{t_{i,n}\leq u}\mathrm{E}_{\sigma_0^2}\left\{|U_i^n|^2 I_{\left\{|U_i^n|>\varepsilon\right\}}\middle|\mathcal{F}_{t_{i-1,n}}\right\} \qquad (2.B.10)\\
&\leq \sum_{i=1}^n \frac{n\,(t_{i,n}-t_{i-1,n})^2+1/n}{M^2}\mathrm{E}\left\{\left(Z_{i,n}^2-1\right)^2 I_{\left\{\sqrt{n(t_{i,n}-t_{i-1,n})^2+1/n}\left|Z_{i,n}^2-1\right|>M\varepsilon\right\}}\right\}\\
&\leq \max_{i=1,\dots,n}\mathrm{E}\left\{\left(Z_{i,n}^2-1\right)^2 I_{\left\{\sqrt{n(t_{i,n}-t_{i-1,n})^2+1/n}\left|Z_{i,n}^2-1\right|>M\varepsilon\right\}}\right\}\sum_{i=1}^n\frac{n\,(t_{i,n}-t_{i-1,n})^2+1/n}{M^2},
\end{aligned}
$$

which converges to zero in view of Assumption 2.6 and the fact that the mapping $x \mapsto \mathrm{E}\left\{\left(Z_{i,n}^2-1\right)^2 I_{\left\{\sqrt{x}\left|Z_{i,n}^2-1\right|>M\varepsilon\right\}}\right\}$ is continuous at $x = 0$.

Finally, the quadratic variation of the limiting Gaussian process follows from the law of large numbers by

$$
\begin{aligned}
&\sum_{t_{i,n}\le u} \mathrm{E}_{\sigma_0^2}\left\{U_i^n U_i^{n\prime} | \mathcal{F}_{t_{i-1,n}}\right\} \qquad (2.B.11)\\
&= \quad 2\sum_{t_{i,n}\le u} n \begin{bmatrix} 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \\ [n(t_{i,n}-t_{i-1,n})]^{-1} & [n(t_{i,n}-t_{i-1,n})]^{-2} \end{bmatrix} \left(\frac{(t_{i,n}-t_{i-1,n})^2}{\int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v}\right)^2 \\
&= \quad 2\sum_{t_{i,n}\le u} \begin{bmatrix} n(t_{i,n}-t_{i-1,n}) & 1 \\ 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \end{bmatrix} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}v \\
&\qquad +2\sum_{t_{i,n}\le u} \begin{bmatrix} n(t_{i,n}-t_{i-1,n}) & 1 \\ 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \end{bmatrix} (t_{i,n}-t_{i-1,n}) \\
&\qquad\qquad \times \left(\left[\frac{1}{t_{i,n}-t_{i-1,n}} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v\right]^{-2} - \frac{1}{t_{i,n}-t_{i-1,n}} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}v\right) \\
&= \quad 2\sum_{t_{i,n}\le u} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}\begin{bmatrix} T_n^{-1}(v) & v \\ v & T_n(v) \end{bmatrix} + o_P(1) \\
&\to \quad 2\int_{v=0}^{u} \sigma_0^4(v)\mathrm{d}\begin{bmatrix} T^{-1}(v) & v \\ v & T(v) \end{bmatrix},
\end{aligned}
$$

where the $o_P(1)$-term follows from (2.A.7) combined with Assumption 2.5 and where the final convergence follows from the weak convergence condition in Assumption 2.6.

2.B.3 Proof of Theorem 2.2.

The proof consists of showing that the likelihood ratio satisfies the appropriate quadratic expansion. First, observe

$$\frac{t_{i,n}-t_{i-1,n}}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} = \frac{t_{i,n}-t_{i-1,n}}{\sigma_0^2(t_{i-1,n},t_{i,n})} + \frac{\alpha}{\sqrt{n}} \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u)h(u)\mathrm{d}u. \qquad (2.B.12)$$

Now, the log-likelihood ratio of $\mathbb{P}_\alpha^{(n)}$ with respect to $\mathbb{P}_0^{(n)}$ is, in view of Assumption 2.3 and 2.4, given by

$$
\begin{aligned}
\log \frac{\mathrm{d}\mathbb{P}_\alpha^{(n)}}{\mathrm{d}\mathbb{P}_0^{(n)}} &= \frac{1}{2}\sum_{i=1}^n \log \frac{\sigma_0^2(t_{i-1,n},t_{i,n})}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} \\
&+\frac{1}{2}\sum_{i=1}^n \frac{\left[R_{i,n} - \left[\mu(t_{i-1,n},t_{i,n}) + \gamma(t_{i-1,n},t_{i,n})\sigma_0^2(t_{i-1,n},t_{i,n})\right](t_{i,n}-t_{i-1,n})\right]^2}{\sigma_0^2(t_{i-1,n},t_{i,n})\,(t_{i,n}-t_{i-1,n})} \\
&-\frac{1}{2}\sum_{i=1}^n \frac{\left[R_{i,n} - \left[\mu(t_{i-1,n},t_{i,n}) + \gamma(t_{i-1,n},t_{i,n})\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\right](t_{i,n}-t_{i-1,n})\right]^2}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\,(t_{i,n}-t_{i-1,n})} \\
&= \frac{1}{2}\sum_{i=1}^n \log \frac{\sigma_0^2(t_{i-1,n},t_{i,n})}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} \\
&-\frac{1}{2}\sum_{i=1}^n \frac{R_{i,n}^2}{t_{i,n}-t_{i-1,n}}\left[\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} - \frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}\right] \\
&+\sum_{i=1}^n R_{i,n}\mu(t_{i-1,n},t_{i,n})\left[\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} - \frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}\right] \\
&+\frac{1}{2}\sum_{i=1}^n \mu(t_{i-1,n},t_{i,n})^2\,(t_{i,n}-t_{i-1,n})\left[\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})} - \frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}\right] \\
&+\frac{1}{2}\sum_{i=1}^n \gamma(t_{i-1,n},t_{i,n})^2\,(t_{i,n}-t_{i-1,n})\left[\sigma_0^2(t_{i-1,n},t_{i,n}) - \sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\right] \\
&= \frac{1}{2}\sum_{i=1}^n \log\left[1+\frac{\alpha}{\sqrt{n}}\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right] \\
&-\frac{\alpha}{2\sqrt{n}}\sum_{i=1}^n \frac{R_{i,n}^2}{(t_{i,n}-t_{i-1,n})^2}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u + u_n \\
&= -\frac{\alpha}{2}\int_0^1 \sigma_0^{-2}(u)h(u)\mathrm{d}\sqrt{n}\left[RV_n^*(u) - \int_{v=0}^u \sigma_0^2(v)\mathrm{d}T_n(v)\right] \\
&-\frac{\alpha^2}{4}\int_{u=0}^1 h^2(u)\mathrm{d}T(u) + u_n + r_n,
\end{aligned}
\tag{2.B.13}
$$

with

$$
\begin{aligned}
u_n &= \sum_{i=1}^{n} R_{i,n}\mu(t_{i-1,n},t_{i,n})\left[\frac{1}{\sigma^2_{\alpha/\sqrt{n}}(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma^2_0(t_{i-1,n},t_{i,n})}\right]\\
&\quad +\frac{1}{2}\sum_{i=1}^{n}\mu(t_{i-1,n},t_{i,n})^2\left(t_{i,n}-t_{i-1,n}\right)\left[\frac{1}{\sigma^2_0(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma^2_{\alpha/\sqrt{n}}(t_{i-1,n},t_{i,n})}\right]\\
&\quad +\frac{1}{2}\sum_{i=1}^{n}\gamma(t_{i-1,n},t_{i,n})^2\left(t_{i,n}-t_{i-1,n}\right)\left[\sigma^2_0(t_{i-1,n},t_{i,n})-\sigma^2_{\alpha/\sqrt{n}}(t_{i-1,n},t_{i,n})\right],
\end{aligned}
$$

and

$$
\begin{aligned}
r_n &= \frac{1}{2}\sum_{i=1}^{n}\log\left[1+\frac{\alpha}{\sqrt{n}}\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right] \qquad (2.B.14)\\
&\quad -\frac{\alpha}{2}\sqrt{n}\int_0^1 h(u)\mathrm{d}T_n(u)+\frac{\alpha^2}{4}\int_{u=0}^{1}h^2(u)\mathrm{d}T(u).
\end{aligned}
$$

We need to show that each term in u_n and r_n converges to zero.

The first term of u_n can be bounded, using the Cauchy-Schwarz inequality, by

$$
\begin{aligned}
&\left(\frac{1}{\sqrt{n}}\left[\sqrt{n}\sum_{i=1}^{n}\frac{R^2_{i,n}}{n\left(t_{i,n}-t_{i-1,n}\right)}\right]\right)^{1/2}\\
&\times\left(\sum_{i=1}^{n}\mu(t_{i-1,n},t_{i,n})^2\alpha^2\left[\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\left(t_{i,n}-t_{i-1,n}\right)}\right]^2\left(t_{i,n}-t_{i-1,n}\right)\right)^{1/2},
\end{aligned}
$$

which converges to zero as $\sqrt{n}\sum_{i=1}^{n}R^2_{i,n}/\left(n\left(t_{i,n}-t_{i-1,n}\right)\right)$ is $O_P(1)$ by Lemma 2.1 and μ and σ_0^{-2} are bounded.

The (absolute value of the) second term of u_n can be bounded by

$$
\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\left|\mu(t_{i-1,n},t_{i,n})^2\alpha\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u\right|
$$

which converges to zero as μ is bounded and $\sum_{i=1}^{n}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u=\int_0^1\sigma_0^{-2}(u)h(u)\mathrm{d}u$.

The third term of u_n can be bounded using $|x^{-1}-y^{-1}|\le c^{-2}|x-y|$ for $x\ge c$, $y\ge c$, and $c>0$. Indeed, for n sufficiently large, both $\sigma^2_0(t_{i-1,n},t_{i,n})$ and $\sigma^2_{\alpha/\sqrt{n}}(t_{i-1,n},t_{i,n})$

are larger than $c = \frac{1}{2}\min_{0\leq u\leq 1}\sigma_0^2(u)$. Thus, we can bound the third term of u_n by

$$\frac{1}{2}c^{-2}\sum_{i=1}^{n}\gamma(t_{i-1,n},t_{i,n})^2\,(t_{i,n}-t_{i-1,n})\left|\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma^2_{\alpha/\sqrt{n}}(t_{i-1,n},t_{i,n})}\right|,$$

and we may proceed as for the second term of u_n.

Now, let's concentrate on r_n. Using $\|h\|\leq 1$ and $|\log(1+x)-x+x^2/2|<|x|^3$ for $|x|<1/2$, we have, for $n\geq 4$,

$$\begin{aligned}|r_n| \leq\ & \frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}h(u)\mathrm{d}u\right| \qquad (2.B.15)\\ &+\frac{\alpha^2}{4n}\sum_{i=1}^{n}\left|\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right)^2-n\int_{t_{i-1,n}}^{t_{i,n}}h^2(u)\mathrm{d}T_n(u)\right|\\ &+\frac{\alpha^2}{4}\left|\int_0^1 h^2(u)\mathrm{d}\left[T_n-T\right](u)\right|\\ &+\frac{\alpha^3}{2n\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right|^3.\end{aligned}$$

We need to show that each of these four remainder terms converge to zero. For the last term, this is obvious as each element in the sum is bounded by 1 since $\|h\|\leq 1$. Convergence of the third term follows from the weak convergence in Assumption 2.6.

The second term can be bounded, using (2.A.8), as

$$\begin{aligned}
&\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\left|\left((t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u\right)^2\right.\\
&\qquad\left.-\left((t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u\right)^2(t_{i,n}-t_{i-1,n})^{-1}\int_{t_{i-1,n}}^{t_{i,n}}h^2(u)\mathrm{d}u\right|\\
&\leq\ \frac{\alpha^2}{n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\\
&\quad+\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)\\
&\leq\ \alpha^2\left\|\sigma_0^2\right\|^3\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)}\\
&\quad+\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right),
\end{aligned}$$

which converges to zero by Assumption 2.3 and since

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)\leq\max_{i=1,\ldots,n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)\to 0, \tag{2.B.16}$$

due to the fact that h is cadlag on $[0,1]$ and $\max_{i=1,\ldots,n}(t_{i,n}-t_{i-1,n})\to 0$.

Finally, for the first term in (2.B.15), observe

$$\begin{aligned}
&\frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}h(u)\mathrm{d}u\right|\\
&=\ \frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\left[\sigma_0^{-2}(u)-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v\right]h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right|\\
&\leq\ \frac{\alpha}{2\sqrt{n}}\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\\
&\leq\ \left[\max_{i=1,\ldots,n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\right]\frac{\alpha}{2}\left\|\sigma_0^2\right\|\left[\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)\right]^{1/2}\to 0,
\end{aligned}$$

in view of (2.B.16).

To complete the proof of this LAN result, observe, using Lemma 2.1,

$$
\begin{aligned}
&-\frac{1}{2}\int_{u=0}^{1}\sigma_0^{-2}(u)h(u)\mathrm{d}\sqrt{n}\left[RV_n^*(u)-\int_{v=0}^{u}\sigma_0^2(v)\mathrm{d}T_n(v)\right] \\
&\to_L \quad -\frac{1}{\sqrt{2}}\int_{u=0}^{1}h(u)\sqrt{T'(u)}\mathrm{d}W(u)\sim N\left(0,\frac{1}{2}\int_{u=0}^{1}h^2(u)\mathrm{d}T(u)\right).
\end{aligned}
$$

The stated contiguity is a well-known consequence of Le Cam's first lemma, see, e.g., van der Vaart (2000), Lemma 6.4.

2.B.4 Proof of Theorem 2.4.

We start this proof with a lemma that describes the limiting behavior of the derivative of the smoothed realized variance RV_n^S.

Lemma 2.10. *Under Assumption 2.8, we have*

$$
RV_n^{S\prime}(u)=K'(u-1)RV_n^*(1)+\int_{v=0}^{1}K''(u-v)RV_n^*(v)\,dv. \tag{2.B.17}
$$

Moreover, under Condition 1 and Assumption 2.7–2.8, we have the following convergence of the process $RV_n^{S\prime}$, under $\mathbb{P}_{\sigma_0^2}^{(n)}$,

$$
\begin{aligned}
&\sqrt{n}\left[RV_n^{S\prime}(u)-\int_{w=0}^{1}K'(u-w)\sigma_0^2(w)\,dT_n(w)\right] \\
&\to_L \quad \sqrt{2}\int_{w=0}^{1}K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\,dW(w).
\end{aligned}
\tag{2.B.18}
$$

Proof. n Relation (2.B.17) follows directly from partial integration and $RV_n^*(0)=0$. Relation (2.B.18) follows by applying the Continuous Mapping theorem to $f\mapsto K'(\cdot-1)f(1)+\int_{v=0}^{1}K''(\cdot-v)f(v)\mathrm{d}v$ (which is linear and bounded in view of the boundedness of both K' and K''). More precisely, from Condition 1[9],

$$
\begin{aligned}
&\sqrt{n}\left[RV_n^{S\prime}(u) - \int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right] \\
&= \sqrt{n}\left[RV_n^{S\prime}(u) - K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)\mathrm{d}T_n(v) - \int_{w=0}^{1}\int_{v=w}^{1} K''(u-v)\mathrm{d}v\sigma_0^2(w)\mathrm{d}T_n(w)\right] \\
&= \sqrt{n}\left[RV_n^{S\prime}(u) - K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)\mathrm{d}T_n(v) - \int_{v=0}^{1} K''(u-v)\int_{w=0}^{v}\sigma_0^2(w)\mathrm{d}T_n(w)\mathrm{d}v\right] \\
&\to_L \sqrt{2}K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)T'(v)^{1/2}\mathrm{d}W(v) \\
&\quad + \sqrt{2}\int_{v=0}^{1} K''(u-v)\int_{w=0}^{v}\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w)\mathrm{d}v \\
&= \sqrt{2}K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)T'(v)^{1/2}\mathrm{d}W(v) \\
&\quad + \sqrt{2}\int_{w=0}^{1}\int_{v=w}^{1} K''(u-v)\mathrm{d}v\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w) \\
&= \sqrt{2}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w).
\end{aligned}
$$

■

In view of Lemma 2.10 and the almost sure convergence of T_n' in Assumption 2.6

$$
\begin{aligned}
&\sqrt{n}\left[T_n'(u)^{-1}RV_n^{S\prime}(u) - T_n'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right] \\
&\to_L \sqrt{2}T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w).
\end{aligned}
$$

Hence, applying the delta method for the transformation $(u, x(u)) \mapsto g(u, x(u))$, we find

$$
\begin{aligned}
&\sqrt{n}\left[g\left(u, T_n'(u)^{-1}RV_n^{S\prime}(u)\right) - g\left(u, T_n'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right)\right] \\
&\to_L \sqrt{2}\frac{\partial g}{\partial \sigma^2}\left(u, T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T(w)\right) \\
&\qquad \times T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w),
\end{aligned}
$$

[9] In the remainder of this appendix, we abuse notation by specifying a process using its value at time u. All statements should be read as weak convergence in $D[0,1]$.

Applying the Continuous Mapping theorem to the bounded linear functional $x(\cdot) \mapsto \int_{u=0}^{1} x(u)\mathrm{d}u$ leads to (2.4.26).

2.B.5 Proof of Proposition 2.1

If K converges weakly to the distribution function $I\{\cdot \geq 0\}$, we have that

$$\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}K(u-w),$$

converges pointwise to

$$\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}I(u-w \geq 0) = -T'(u)\sigma_0^2(u).$$

Consequently, from the bounded convergence theorem (recall that T' is bounded away from zero and that T', σ_0^2, g, and K' are bounded), we find that

$$\int_{u=0}^{1} g\left(u, -T'(u)^{-1}\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}K(u-w)\right)\mathrm{d}u$$

converges to

$$\int_{u=0}^{1} g\left(u, \sigma_0^2(u)\right)\mathrm{d}u = \psi_g\left(\sigma_0^2\right).$$

Let $g^{(2)}$ denotes the derivative with respect to the second argument of g, i.e., $g^{(2)}(u,\sigma^2) = \frac{\partial g}{\partial \sigma^2}(u,\sigma^2)$. Concerning the limiting distribution, observe that we can rewrite it as

$$\begin{aligned}
&\sqrt{2}\int_{u=0}^{1} g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right) \\
&\qquad \times T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w)\mathrm{d}u \\
&= \sqrt{2}\int_{w=0}^{1}\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w) \\
&\qquad \times \sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w),
\end{aligned}$$

from which the variance (2.4.29) follows. The convergence of the limiting variance when K converges weakly to point mass at zero, follows as above. In particular, we have

$$\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v) \to -\sigma_0^2(u)T'(u),$$

pointwise in u. Hence, we also have the pointwise convergence

$$g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right) \to g^{(2)}\left(u, \sigma_0^2(u)\right),$$

Now, observe that

$$\begin{aligned}
&\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w)\\
&\quad - \int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, \sigma_0^2(u)\right)\mathrm{d}I\{u-w \geq 0\}\\
&= \int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}\left[K(u-w) - I\{u-w \geq 0\}\right]\\
&\quad + \int_{u=0}^{1} T'(u)^{-1}\left[g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right) - g^{(2)}\left(u, \sigma_0^2(u)\right)\right]\mathrm{d}I\{u-w \geq 0\},
\end{aligned}$$

converges to zero, pointwise in w. Indeed, this follows as the integrand in the first term on the right-hand side is bounded (and the weak convergence of K), while for the second term on the right-hand side we can, again, apply the dominated convergence theorem. Thus, we have established, pointwise in w,

$$\begin{aligned}
&\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w)\\
&\to \int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, \sigma_0^2(u)\right)\mathrm{d}I\{u-w \geq 0\}\\
&= T'(w)^{-1}g^{(2)}\left(w, \sigma_0^2(w)\right),
\end{aligned}$$

as K converges weakly to pointmass at zero. Now, the convergence in (2.4.29) follows from another application of the bounded convergence theorem.

Chapter 3

Monetary Policy Risk in the Cross–Section of Expected Returns

Abstract

This paper examines the pricing implications of monetary policy actions by the Federal Open Market Committee in the cross–section of stock returns. We find that unanticipated changes to the Fed funds target rate is a priced risk factor in the cross–section of S&P 500 constituents, and these carry a significant negative price of risk. This translates into stocks which are positively (negatively) exposed to the monetary policy shocks earning lower (higher) average returns. The results hold in the presence of the market factor, and estimates are remarkably similar when using individual stocks or portfolios as test assets.

3.1 Introduction

The goals of US monetary policy are defined in terms of macroeconomic aggregates, in particular price stability, maximum employment and output. The policy maker, here the Federal Reserve, takes actions through instruments which are at best indirectly geared towards achieving those goals. Bernanke and Kuttner (2005) further state that *"by affecting asset prices and returns, policy makers try to modify economic behaviour in ways that will help to achieve their ultimate objectives."* The naturally arising challenge is to resolve the form of connecting links, if any, between these three variables 1.) policy making decisions, 2.) asset prices, and 3.) economic activity.[10]

Considerable interest lies in understanding the links between asset prices and monetary policy with a focus on the time series relations between actions undertaken by the Federal Reserve and asset returns in fixed income, foreign exchange, and aggregate equity markets.[11] The literature documents consistent and sizeable effects of monetary policy actions onto these asset classes. From an arbitrage pricing theory perspective of Ross (1976a), such findings suggest to consider monetary policy actions as a potential systematic (common) risk factor. The confirmed significance of monetary policy shocks having an effect on expected future market returns in the time series dimension, as in Bernanke and Kuttner (2005), also points into this direction. However, surprisingly, the question of how monetary policy shocks are related to the cross–section of expected returns has received less attention so far. If monetary policy shocks are a systematic risk factor, then, within an arbitrage pricing theory setting or a linear–factor model setting, the cross–section of expected

[10]A rapidly growing body of research analysis the links between asset prices and economic activity. Bloom (2009) focusses on the role of stochastic volatility of stock market returns on the economy (investment, employment, output) with a particular focus on recessions; also see Gilchrist et al. (2011), Stock and Watson (2012), Christiano et al. (forthcoming), Bloom et al. (2014) among others.

[11]See, among others, Kuttner (2001), Andersen et al. (2003), Rigobon and Sack (2004), Bernanke and Kuttner (2005), and Gürkaynak et al. (2005).

stock returns must, at least partly, be explained by their respective sensitivities to these shocks.

In this study we seek to understand if monetary policy risks are priced in the cross–section of stocks, and to estimate the price of monetary policy risk. We first provide an investigation of the patterns between average returns and monetary policy shocks through a portfolio sorting methodology. We document a remarkable pattern of declining average excess returns of the sorted portfolios with increasing exposure to monetary policy shocks for given levels of exposures to market risk on FOMC announcement days. Moreover, the average returns of our portfolios on these days are remarkably larger compared to non–announcement days.

As it is unlikely that stock prices respond to anticipated information about policy actions, we define monetary policy shocks as the "surprise" component in target rate changes. In order to estimate exposures of individual stock returns to factors as precisely as possible, we make use of intraday data. In particular, we use intraday event windows around the FOMC press releases to measure the response of individual stock prices to monetary policy shocks.[12]

We find that shocks to monetary policy carry a statistically significant negative price of risk on FOMC announcement days. This translates to stocks which are positively (negatively) exposed to monetary policy shocks earning lower (higher) average returns, all else being equal; thus, supporting the empirical patterns across constructed portfolios. The results are in line with economic reasoning: Monetary tightening is usually associated with "bad news" such as high inflationary expectations. Assets which do well with arrival of such news would be desirable for the risk averse investor to pay a premium for holding them, and hence lowering their average returns.

Moreover, we analyze the prices of risk for the market and monetary policy shocks

[12]Using intra–day data for the purpose of obtaining precise estimates of the bond price exposures in the context of macroeconomic news announcements is not a new tradition. See, e.g., Ederington and Lee (2001), Balduzzi et al. (2001)

at the intraday level. In particular, we obtain Fama–Macbeth risk price estimates for both the market and the monetary policy shocks for three distinct intraday observation periods on announcement days: the pre–announcement window (PAW), the announcement window (AW), and the post–announcement window (POAW), together spanning the cash market opening hours from 9:30 to 16:00 EST. We find that the largest proportion of monetary policy risk premiums are earned during the pre–announcement window. In particular, consistent with the evidence based on daily observations, we find that monetary policy shocks carry a statistically significant and negative risk premium whereas the market carries a statistically significant positive risk premium during the pre–announcement window. However, the prices of risk for the monetary policy shocks are not significant for both announcement windows and post–announcement windows.

This chapter contributes to the literature on how monetary policy affects asset prices. The impact of unexpected changes in Fed funds target rate on long term interest rates and on the aggregate equity market has previously been studied by Kuttner (2001)[13] and Bernanke and Kuttner (2005)[14] using daily event windows. Gürkaynak et al. (2005) carry out similar analyses at the intraday level for S&P 500 index returns only.

This chapter is also related to the growing literature regarding the behaviour of asset prices around scheduled macroeconomic announcements. De Goeij et al. (2013) shed light on a variety of non–monetary macroeconomic announcement news and their pricing in the cross–section of stock returns. We here focus on the case of FOMC announcements which might reveal forward looking information about financial markets and the state of the economy as suggested by Romer and Romer (2000), which makes them interesting from an asset pricing perspective. Savor and Wilson (2013) attribute more than half of the equity market risk premium to only

[13] A 100 basis point unexpected increase would lead to a 30 basis points increase in ten year interest rates.

[14] A 25 basis point surprise cut in Fed funds target rate would lead to a 100 basis points increase in the CRSP value weighted index.

those days on which inflation, unemployment and FOMC announcements occur.[15] Cieslak et al. (2015) show an apparent biweekly excess return cycle following FOMC announcements.[16].

The rest of the paper is organized as follows. In Section 3.2 we outline the research design. In Section 3.3 we present our analysis, first describing the proxy for monetary policy shocks and the cross–sectional stock return data. Then, we move on to the empirical results in Section 3.3.3. Section 3.5 concludes. The Appendix entails a proof, our data cleaning procedures, followed by the Tables.

3.2 Pricing Monetary Policy Shocks in the Cross–Section

3.2.1 Model Specification

In the absence of arbitrage, there exists a stochastic discount factor which prices any traded asset. Linear factor models additionally specify the stochastic discount factor to be in a linear form. In order to assess whether exposures to monetary policy risk are priced in the cross–section of stocks, we will make use of this framework and specify the cross–section of expected returns in the following way: Let the expected return on stock i, denoted by $\mathrm{E}\left[R_t^{i,e}\right]$, take the form

$$\mathrm{E}\left[R_t^{i,e}\right] = a + \beta^i_{MKT}\lambda_{MKT} + \beta^i_S\lambda_S, \tag{3.2.1}$$

where a is a constant, β^i_{MKT} is the loading of asset i on the excess market return, and β^i_S is asset i's sensitivity to monetary policy risk. The price of risk of the market factor and the price of monetary policy risk are denoted by λ_{MKT} and λ_S

[15] The results in Savor and Wilson (2013) suggests that 60% of annual equity risk premium is earned on days with inflation, unemployment, or FOMC target rate announcements.

[16] Their study suggests that aggregate excess market returns exhibit a bi–weekly pattern such that the entire equity premium has been earned in even weeks in FOMC cycle time.

respectively. Note that if a factor is traded, i.e. excess return, its price of risk is equal to the expected return of the factor.

Eq. (3.2.1) links the expected return of asset i to its exposure to the market risk, β^i_{MKT}, and its exposure to monetary policy risk, β^i_S. The main implication from the factor model setting in 3.2.1 is that assets with different factor loadings on monetary policy shocks have different expected returns while controlling for market risk. While the true news arrival process related to monetary policy may not be restricted to scheduled and unscheduled FOMC announcements only, we focus on the "FOMC day" events in our empirical analysis—since on these days there are for sure observations available to the researcher with an exact release time stamp.

Going further, we want to estimate the exposures to the risk factors as precisely as possible and we can achieve this by making use of high frequency data. That is, we estimate factor exposures in narrow windows surrounding the announcements, to reduce the impact of unrelated return variation "noise"[17], in the realized price paths of assets. The proof in Appendix 3.A.1 details why this is the case in a simple setting. Empirically, this is in line with the findings of Gürkaynak et al. (2005) at the index level in that standard errors of coefficient estimates, from an OLS regression of market returns on fed funds surprise target changes, are increasing and R squares are decreasing when going from a narrow 30–minute intraday to a daily event window.

Moreover, Rigobon and Sack (2004) and Gürkaynak and Wright (2013) point out that using intraday data helps circumventing the otherwise present omitted variable bias and inconsistencies caused by simultaneity in longer horizons when using e.g. daily or monthly data. For illustrative purposes, consider the FOMC target rate announcement on November, 15$^{\text{th}}$ 1995 which coincides with news releases regarding consumer price index, industrial production, capacity utilization, and business inventories earlier on the same day—and all of which presumably entered the decision

[17]These include but are not limited to the release of other firm specific announcements, e.g. earnings announcements or analyst report releases, and other macroeconomic announcements such as GDP news, industrial production news, employment reports among many others.

making process of the FOMC board members. In such cases, using daily returns will cause estimates to be inconsistent in estimating the true impact of monetary policy decisions onto asset prices. Using narrow intraday windows surrounding the FOMC press release time thus decreases the possibility of existence of a joint response of monetary policy and asset prices to other news (omitted variable bias), and decreases the possibility of monetary policy responding to movements in asset prices (addresses simultaneity).

Motivated by these arguments, we will use the following specification for estimating the assets' exposures to risk factors

$$R^i_{\tau_d-h,\tau_d+H} = \alpha^i + \beta^i_{MKT} R^{MKT}_{\tau_d-h,\tau_d+H} + \beta^i_S S_{\tau_d-h,\tau_d+H} + \varepsilon^i_{\tau_d-h,\tau_d+H}, \tag{3.2.2}$$

where $R^i_{\tau_d-h,\tau_d+H}$ and $R^{MKT}_{\tau_d-h,\tau_d+H}$ are the excess returns on stock i and the market over the interval of length $[H-h]$ around the announcement at time τ_d on day $d \in D$, which is the set of the FOMC announcement days. S_{τ_d-h,τ_d+H} stands for the monetary policy shock, and β^i_{MKT} and β^i_S are the loadings on market risk and monetary policy risk respectively. Moreover, $\mathrm{E}\left[\varepsilon^i_{\tau_d-h,\tau_d+H}\right] = 0$, $\mathrm{E}\left[\varepsilon^i_{\tau_d-h,\tau_d+H} | S_{\tau_d-h,\tau_d+H}\right] = 0$ and $\mathrm{E}\left[\varepsilon^i_{\tau_d-h,\tau_d+H} | R^{MKT}_{\tau_d-h,\tau_d+H}\right] = 0$. We, throughout, make the assumption that these betas also apply outside the announcement windows, i.e., exposures are constant in the data generating process. This allows us to measure the prices of risk earned on different event windows such as the daily level.[18]

In our empirical analysis, we base our estimation of exposures, β^i_{MKT} and β^i_S, on 30 minute windows (10 minutes before to 20 minutes after) surrounding the annnouncements, that is h stands for 10 and H stands for 20. We also employ 1 hour windows (15 minutes before to 45 minutes after) surrounding the announcements as

[18] News arrival times related to monetary policy may not be restricted to announcement times only, see e.g. Lucca and Moench (2015) and Cieslak et al. (2015). With such approach, we can analyze the premiums earned also on periods where we do not have observable proxies, e.g announcement days, non-announcement windows, FOMC cycles. Findings in this paper confirm that monetary policy shocks earn statistically significant premia on announcement days.

a robustness check. Remember that under the assumption that exposures for the market and monetary policy shocks are constant over the announcement days, these betas apply outside of the event windows. Accordingly, we can obtain the price of risk estimates for announcement day by employing these betas in the second step of Fama–Macbeth (1973) regressions.

3.3 Empirical Analysis

In order to assess whether monetary policy shocks are priced in the cross–section of stocks returns given Eq. (3.2.1), two proxies need to be defined: One for the market factor, which does not leave much choice other than instruments tracking the level of the S&P 500 when the focus shall be on a liquid traded asset, and one for the monetary policy factor, which has to be constructed carefully as it is crucial to accurately measure the unanticipated, exogenous component of monetary policy changes as we shall see in the next section.

3.3.1 Measuring Monetary Policy Shocks: The MPS Factor Proxy

We expect the market subsumes all available information regarding the outcome of monetary policy decisions, and these expectations are incorporated into asset prices. Thus, it is unlikely that the changes that were already expected would move the asset prices. The task at hand is to disentangle the unanticipated component of monetary policy actions from the anticipated ones which might likely be factored into prices.

Kuttner (2001), in line with such reasoning, argues that failure to capture the effects of federal funds target changes on asset prices is due to not disentangling these changes into their the expected and unexpected components.[19] Accordingly,

[19]Fatum and Scholnick (2008) confirm the previously documented failure to capture effects of Fed funds target rate changes in the foreign exchange market is due to not properly isolating unexpected from expected changes in the monetary policy.

he proposed a technique that is based on fed funds futures to identify the unexpected component of the Fed funds target rate using daily event windows. However, Bernanke and Kuttner (2005) point out that using daily measures does not fully solve the simultaneity and omitted variable bias. The Fed funds target rate is likely to be affected by other news within the daily event window, most severely by other macroeconomic announcements which sometimes coinciding with FOMC announcement days.

To avoid these pitfalls, the measure used in this analysis is based on the technique of Gürkaynak et al. (2005) and Fleming and Piazzesi (2005), more recently used by Gorodnichenko and Weber (2014), in modifying the Kuttner decomposition by relying on high frequency (intraday) fed funds futures data. This measure relies on calculating the policy surprise component by examining the changes in the Fed funds futures rate within a narrow window around the FOMC announcement release time such as either a tight 30–minute (TW) or a larger one–hour window (LW), covering the period from 10 minutes before the announcement to 20 minutes after, from 15 minutes before the announcement to 45 minutes after, respectively.

The surprise component is given by

$$S_{\tau_d - h, \tau_d + H} = \frac{P}{P-d} \left(f_{\tau_d + H} - f_{\tau_d - h} \right) \tag{3.3.1}$$

where $[\tau_d - h, \tau_d + H]$ represents the interval surrounding the announcement news release time occurring at τ_d on day d of the month when there are P days in the respective month. f stands for the fed funds futures rate. The term $\frac{P}{P-d}$ adjusts for the fact that the Fed funds futures settles on the average Fed Funds rate for the contract month.

The use of high–frequency returns is particularly important to achieve better identification given that financial markets respond to the FOMC announcements within minutes; see Andersen et al. (2003) and Andersen et al. (2007) although their interest is associated with detecting price discontinuities produced by macroeconomic announcements. In this sense, the measure here differs from Kuttner (2001) and Bernanke and Kuttner (2005) whose surprise measure is more noisy as it is based on daily data. The choice of a coarse relative to highest available (one second) sampling frequency compromises the desire to reduce measurement error and the need to avoid microstructure noise biases and non–synchronicity effects arising at very high frequency.

3.3.2 Cross–Sectional Data and The Market Factor Proxy

Our cross–sectional dataset comprises of all stocks that were throughout constituent of the S&P 500 index during the full sample period of our analysis from January 1, 1995 to December 31, 2009. Relying on firms which were constituents throughout this 15 year period is a tradeoff between choosing the maximum number of FOMC day observations (129) and more firm–FOMC day observations. We accept that there is currently some survivorship bias in our dataset, which is why we aim to extend our results to cover all stocks which were at some point in time constituent of the index. This would double or quadruple the number of assets up to 916 firms depending on the cut–off level, which is quantified in terms of the minimum number of FOMC announcement exposures, e.g. a times series of 30 FOMC announcements corresponds to roughly four years of data, which brings it own caveats along with more severe microstructure noise issues for less liquid stocks.

Consequently, we consider the S&P 500 index as a proxy for the market return. In particular, we use high frequency price data obtained from the Trade and Quote (TAQ) database for the constituents along with the SPYDR ETF, which is a highly liquid exchange traded fund tracking the S&P 500 index. No–arbitrage ensures that the ETF's price does not deviate from the fundamental value of the underlying index.

The cleaning procedure for TAQ data follows largely Barndorff–Nielsen et al. (2009) in removing erroneous entries and assigning price observations to an equidistant one second time grid from which coarser frequencies can be subsampled. Appendix 3.B.1 provides further details regarding our cleaning procedure.

We supplement the high frequency realized price paths with daily data from the Center for Research in Security Prices (CRSP). This is necessary as the TAQ database only provides raw prices and does not adjust for stock splits or dividend distributions, which would in turn lead to large errors in the computation of close–close and overnight returns. Aligning the two data sources correctly is particularly important as we are interested in precisely measuring exposures in narrow event window and examine the pricing effects on a daily horizon. Different methods to reconcile the two databases are currently being applied in the literature but it turns out that there is only way to ensure a correct match. The reader is referred to Appendix 3.B.2 for further details regarding our reconciliation procedure.

Further, we use the one–month T–Bill return as risk–free rate provided Ibbotson and Associates from Kenneth French's website. The excess return on the S&P 500 is calculated from the value–weighted sum of all arithmetic constituent returns on a particular day less the riskless rate.

3.3.3 Empirical Results

In order to get some preliminary insight regarding average returns on FOMC announcement and non–announcement days, this section presents summary statistics for the S&P 500 index. In Table 3.1, we report the descriptive statistics for the ETF tracking the S&P 500, i.e. close–to–close and open–to–close (based on NYSE cash market trading hours) log excess returns for all 129 FOMC announcement days from January 1, 1995 to December 31, 2009. We also report separate statistics for days subject to surprise cuts (an unexpected decrease in Fed Funds target rate) and surprise tightening (an unexpected increase in Fed Funds target rate). The magnitude of log average excess returns on FOMC announcement days are striking, with 41.5 basis points (bps) close–to–close. Overnight returns have no significant impact, almost the entire return is earned within trading hours. Non–announcement day returns are much smaller in magnitude with insignificant 0.7bps. Moreover, looking at days with surprise cuts, we see an increase in magnitude, with 56.1bps from close–to–close and 66.7bps from open–to–close. The picture on the days with surprise tightening is different, with 28.1bps (16.7bps) from close–to–close (open–to–close). This suggests that monetary policy shocks may constitute priced risk factors for stock returns, well in line with Lucca and Moench (2015) who document large average market returns before news announcements. However, examining the asymmetries regarding surprise cuts and tightening, our results further document that the average returns are much larger in magnitude compared to the average returns on days with surprise tightening.

Table 3.2 reports the descriptive statistics for intraday returns for the ETF tracking the S&P 500 index for the same sample period. On all announcement days, average returns are roughly five basis points in magnitude with a standard deviation about one percent (in absolute terms). However, when examining the surprise cut days and surprise tightening days, the average returns are large in magnitude for both 30–minute (TW) and 60–minute (LW) windows surrounding the announcements with a positive sign for surprise cut days and with a negative sign for surprise tightening days. The lower row in Table 3.2 documents almost perfect correlation of returns within the 30–minute tight window and the larger 60–minute window.

[+++ Insert Table 3.1 and 3.2 around here +++]

Portfolio Sorts

Our interest lies in whether monetary policy shocks are priced in the cross–section of stock returns. Note that a factor model of a linear form implies the existence of patterns between the expected return of an asset and its risk factor loadings. For example, in a one factor setting, the capital asset pricing model (CAPM) implies that assets which are more exposed to market risk should earn higher expected returns than assets which are less exposed to market risk, where market risk represents one common systematic factor. In particular, CAPM predicts a positive relationship between expected returns and the loadings on the market risk in the cross–section.

In a general factor model setting with more factors, a factor model implies that assets with different loadings on a certain risk factor should earn different expected returns when all other predictor variables, i.e. the exposures to other risk factors, are kept fixed. Hence, in our setting (Eq. 3.2.1), if shocks to monetary policy is a systematic risk factor, there should be observable patterns between the stocks' realized average returns and their loadings on monetary policy shocks across given levels of exposures to market risk.

Consequently, firstly, we explore whether we can indeed empirically confirm the existence of such patterns between average returns and loadings on monetary policy shocks. Therefore, we form portfolios from assets grouped by their degree of exposure to each of the two risk factors. Constructing portfolios instead of focussing on each single asset case individually is particularly important as it lowers the standard errors of factor exposures, see e.g. Ang et al. (2010)—which becomes even more important given that our sample size is restricted by the number of FOMC announcements.

We proceed in the following way: In the first step, based on a large cross–section of individual stocks, we form quintile portfolios by sorting stocks in terms of their market exposure, β_{MKT}, from low to high (1–5). In the second step, the stocks within each of the five β_{MKT} quintiles are further sorted into new quintile portfolios based on their monetary policy sensitivity as assessed by β_S.[20] Thus, we obtain 25 $\beta_{MKT} \times \beta_S$ equally weighted portfolios with stocks in the (1,1) quintile sharing the lowest loading on both risk factor and vice versa for the (5,5) quintile portfolio, which contains stocks with the highest monetary policy beta, β_S, while controlling for their market exposures via β_{MKT}. Note that the resulting portfolios as well as individual stocks have wide dispersion on factor loadings[21]

Table 3.3 document the post–formation exposures of the 25 $\beta_{MKT} \times \beta_S$ portfolios to monetary policy shocks, β_S.

[+++ Insert Table 3.3 around here +++]

Post–formation exposures of quintile portfolios to monetary policy shocks monotonically increase when going from low to high within each β_{MKT} quintile. Moreover, these post–formation β_S coefficients of quintile portfolios reveal statistical significance in favour of our hypothesis put forward—with only few exceptions that may be attributed to relatively small sample sizes with regard to the less interesting mid quantile exposure combinations. More importantly, factor loadings of low and high exposure portfolios of monetary policy and market risk exhibit high significance. Consequently, going long in the high β_S and shorting the low β_S quintile portfolios (high minus low) within each of the five β_{MKT} bins generates large exposures to monetary policy shocks with high t–statistics. Note that examining the range of β_S across the 25 portfolios confirms the desired dispersion in terms of their loadings on monetary policy shocks.

In Table 3.4 and 3.5, we present the corresponding summary statistics, i.e. the average daily close–to–close excess returns and their t–statistics, for the 25 constructed β_{MKT} x β_S portfolios.

[+++ Insert Table 3.4 and 3.5 around here +++]

[20] Note that market and monetary policy betas are estimated simultaneously.

[21] Sufficiently disperse exposures increases the power of asset pricing tests, see Ang et al. (2006b), which is important in the subsequent section.

We report the average daily excess returns earned on these portfolios separately for announcement and non–announcement days in the upper and lower panel in the left column of the table, t–statistics are shown in the right column. On announcement days we find remarkable patterns: Average returns are economically large in value, ranging from 0.13% to 0.93% per day, and also statistically significant with the exception of two low market beta portfolios, (Low, 4) and (Low, 5). Non–announcement day portfolio returns are much lower in magnitude and not statistically significant. Moreover, looking at the differences between the average returns between announcement and non–announcement days documented on Table 3.5, we see that they are statistically significant for most of the portfolios.

Further note the consistent declining pattern in average portfolio returns on announcement days in Table 3.4 over the low to high β_S exposure quintile portfolios. In particular, within each β_{MKT} quintile, average returns tend to decrease monotonically in portfolios' exposures to monetary policy shocks. The spread obtained from investing in the high β_S quintile portfolio and selling the low β_S portfolio across all market beta quintile portfolios, β_{MKT}, is negative in the range of -0.15% to -0.46% per day, and all are statistically significant, except for the mid market quintile portfolio. These results provide initial empirical support for the idea to consider exposure to monetary policy shocks as a priced risk factor in the cross–section of stocks. Moreover, the pattern of declining average returns among portfolios with increasing β_S indicates a negative price of risk for being exposed to such shocks. In the following section, we test formally if the monetary policy shocks is a priced risk factor in the cross–section of the stock returns.

The Price of Monetary Policy Risk

Having empirically confirmed our first conjecture of the existence of patterns between average returns and loadings for monetary policy shocks, we now set out to test more formally whether monetary policy shocks constitute a priced risk factor in the cross–section of the stock returns.

We base the analysis on three sets of test assets, the first being the 25 $\beta_{MKT} \times \beta_S$ portfolios, the second set comprises all individual assets, and the third set combines both portfolios and individual assets. The reason behind using individual stocks as test assets is based on the work of Ang et al. (2010), which documents that using individual assets in testing prices of risk factors provides informational gains in estimating of the prices of risk[22]. Moreover, using both the individual stocks and 25 portfolios may provide advantages given the larger number of observations in the cross–section.

In terms of estimation, we apply the standard Fama–Macbeth procedure. In particular, we perform cross–sectional regressions of stock returns on the exposures, $\beta_{MKT} \times \beta_S$, augmented with a constant for all days within the set of FOMC announcement days. This leads to a time series of the coefficients and intercept. The statistical significance is evaluated using Newey–West corrected standard errors. Table 3.6 documents the results with each of the three columns referring to the respective set of assets used.

[+++ Insert Table 3.6 around here +++]

The estimated price of monetary policy risk is -0.066% per announcement day when 25 portfolios are used as test assets in column one. Using individual stocks as test assets instead does not move the estimate of the price of monetary policy risk, then being at -0.067% per announcement day. Moreover, as can be observed in the last column, increasing the cross–section by using both individual stocks and the 25 portfolios does not alter the estimate up to the three decimal places reported. The estimates are all statistically significant with t–statistics which are large in

[22]The efficiency gain is led by the information in individual asset betas. Forming portfolios shrinks the cross–sectional dispersion in the factor exposures and hence leads to informational losses. This in turn translates into lower precision in the prices of risk.

magnitude (3.71, 3.69, and 3.70), while estimates of intercepts are not statistically significant across all three cases. Also note that using the 25 $\beta_{MKT} \times \beta_S$ portfolios as test assets leads to large average adjusted R^2 of 0.32, compared to 0.07 when using either individual stocks only, or using individual stock and portfolios jointly.

The results reported in Table 3.6 are consistent with our hypothesis that monetary policy exposure is priced in the cross–section and it commands a negative risk premium. To gauge the magnitude at an annualized equivalent level, we can multiply the price of risk times the expected intensity which amounts to approx. -0.63% per annum. Recall that the estimates of the prices of risk represent the average premiums earned for a one unit exposure to factor shocks. Revisiting Table 3.3, we can gauge the range of announcement day risk premiums demanded by the 25 $\beta_{MKT} \times \beta_S$ portfolios with a few examples. For instance, the high β_{MKT}, low β_S portfolio, (5, 1), demands an additional premium of 0.33% per day attributable only to its exposure to monetary shock. On the other side, the (3, 5) portfolio has a positive loading on the negative price of monetary policy risk resulting in a reduction of its total risk premium of -0.15% (2.33 $\times$ -0.067%) per day.

3.4 Prices of Risk at the Intraday Level

In previous sections, we estimated the risk prices attached to the market and monetary policy shocks on announcement days and analyzed whether the high average announcement day returns represents any compensation for exposures to monetary policy shocks. Thanks to the high frequency data, we can zoom into the day and analyze the risk prices at different intervals through the day.

We analyze the prices of risk for both the market and the monetary policy shocks for three seperate windows through day from cash market opening at 9:30 ET to closing at 16:00 ET: pre–announcement window (PAW), announcement window (AW), post–announcement window (POAW). Pre–announcement window refers to the period from the cash market opening at 9:30 to 10 minutes before the FOMC announcement, whereas post–announcement window refers to the period from 20 mins after the announcement time to the cash market closing at 16:00. Recall that announcement window covers from 10 minutes before the release of the announcement to 20 minutes thereafter. Panel A of Table 3.6 presents the Fama–Macbeth estimates of risk prices for both the market and the monetary policy shock factor for PAW, AW, POAW. The same methodology as in the previous sections for the estimation of exposures to the market and the monetary policy shocks are applied here, that is the exposures are estimated from the windows surrounding the FOMC announcements, which are useful to gain from precision (we refer the reader to Appendix 3.6 for a proof.). The underlying assumption is that the betas are constant over announcement days, hence such betas apply outside of the announcement windows. The sample covers the same period as the previous sections cover, from January 1, 1995 to December 31, 2009.

The first three rows on Table 3.6 are the Fama–Macbeth risk price estimates. λ_M and λ_S stands for the price of risk for the market, and the price of monetary policy shocks for announcement window respectively. The second three rows are the Fama–Macbeth results for the post–announcement windows and last three rows presents the results for the pre–announcement windows. Looking at the results for the pre-announcement windows, we see that the prices of risk for both the market and the monetary policy shocks are both economically and statistically significant. In particular, we find negative price of risk for the monetary policy shock with magnitude of 0.033%, whereas the price of risk for the market is positive with 0.383%. However, the estimates for both the market and the monetary policy risk prices are both insignificant for post–announcement windows. Moreover, market pries of risk is significant however negative for announcement windows, although the price of risk for monetary policy shocks is not significant.

3.5 Conclusions

We set out to answer the question whether monetary policy actions constitute a priced risk factor, and we indeed confirm that monetary policy shocks are priced in the cross–section of stocks by analyzing highly liquid S&P 500 constituents. In particular, we find that stocks with higher exposures to monetary policy shocks earn lower average announcement day returns. The price attached to monetary policy shocks turns out to be negative. The negative price of monetary policy risk, λ_S, is statistically significant, and robust to the presence of the market factor. The estimated magnitude of λ_S is remarkably similar when using individual stocks, constructed portfolios, or the combined set as test assets.

Examining the prices of risk at the intraday level shows that most of the monetary policy risk premium and the market risk premium is earned during the preannouncement windows. This result is in line with Lucca and Moench (2015) who provide the cumulative return for the market only. The lead lag order relationships and potential volatility feedback effects at work could be further examined, however, there is not a well establish theory for these effects yet and inference might be plagued by several arising biases that occur at high frequency causalities, see AitShalia (2013).

The negative sign of the price of risk is consistent within the setting of Merton's (1973) ICAPM. Bernanke and Kuttner (2005) mention that monetary tightening could reduce the expected level of consumption or monetary policy tightening is usually associated with high inflationary expectations.

Thus, the paper highlights, firstly, the identification of the surprise component of monetary policy changes, i.e. the surprise component of changes in Fed funds target rate, as a risk factor, and, secondly, the use of cross–section of high frequency returns in order to identify the effects of monetary policy shocks more accurately. Thirdly, the analysis shows that some assets have large positive exposure to monetary policy shocks, while others have large negative exposures. One research direction would be to further investigate into the determinants of these sensitivities by incorporating individual firm–level characteristics such as financial leverage, financing cash flows, Tobin's Q ratio. Moreover, to understand the fundamental mechanisms between the cross–section of returns and monetary policy actions, these patterns need to be explained within an asset pricing model, which we intend to address in future work.

3.6 Tables

Table 3.1: Summary Statistics of S&P 500 market returns on announcement days

This table presents the summary statistics of the average log excess returns on the ETF tracking S&P 500 index for all FOMC event days, surprise cut and surprise tightening from January 1, 1995 to December 31, 2009 (in percentages), where c–c denotes close–to–close returns and o–c denotes open–to–close returns. N–aday denotes non–FOMC announcement days for comparison.

in %	**FOMC Days**		**Surprise Cuts**		**Surprise Tightening**		**Non–A Days**
	c–c	o–c	c–c	o–c	c–c	o–c	c–c
Mean	0.415	0.412	0.561	0.667	0.281	0.167	0.007
Std.	1.448	1.300	1.649	1.449	1.453	1.297	1.230
Max.	5.105	5.056	5.105	5.056	4.104	3.887	10.632
Min.	-5.185	-3.164	-5.185	-2.072	-2.654	-3.164	-9.605
# Obs.	129	129	60	60	39	39	3587

Table 3.2: Summary Statistics of S&P 500 market returns on announcement days

This table presents the summary statistics of ETF tacking the S&P 500 index provided by for all FOMC event days between January 1, 1995 to December 31, 2009 and also for the days of suprise cuts (a surprise decrease) and surprise tightening (a surprise increase). TW stands for 30 minutes window surrounding the announcements, LW stands for 60 minutes window surrounding the announcements.

in %	**FOMC Days**		**Surprise Cuts**		**Surprise Tightening**		**Non–A Days**	
	TW	LW	TW	LW	TW	LW	TW	LW
Mean	-0.047	0.058	0.111	0.231	-0.228	-0.165	0.000	0.000
Std.	1.004	1.021	1.303	1.296	0.750	0.793	0.264	0.375
Max.	4.260	3.391	4.260	3.391	2.356	1.978	2.294	3.984
Min.	-6.420	-5.502	-6.420	-5.502	-1.806	-1.856	-1.647	-2.633
# Obs.	129	129	60	60	39	39	3587	3587
Corr.	0.904							

Table 3.3: Betas

This table reports the post–formation β_S of 25 β_{MKT} x β_S equally weighted portfolios. Portfolios are constructed in the following way: exposures of each individual stock, β_{MKT} and β_S, are obtained time series regression with high frequency data as in equation (3.2.2). Stocks are assigned into five quintiles according to their β_{MKT} from lowest (quintile 1) to highest (quintile 5). Then, the stocks within each of the five β_{MKT} quintiles are further sorted into new quintile portfolios based on their monetary policy sensitivity β_S. The row "H–L" refers to the difference in daily returns between the two extreme β_S portfolios. The right block report t-statistics. Sample period covers the period from January 1, 1995 to December 31, 2009.

	B_S quintiles											
β_M quintiles	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
	B_S						**t-statistics for B_S**					
Low	-3.87	-0.73	-0.20	0.63	1.50	5.38	-4.73	-1.48	-0.42	1.64	4.41	8.27
2	-1.72	-0.68	0.31	1.10	2.04	3.76	-4.28	-1.72	0.95	3.07	4.66	7.86
3	-2.44	-1.26	-0.41	0.74	2.33	4.77	-7.80	-4.35	-1.21	2.45	7.97	12.04
4	-3.60	-1.57	-0.52	0.46	1.95	5.54	-8.60	-6.25	-2.20	1.32	5.58	12.34
High	-4.90	-2.51	-0.29	1.85	8.00	12.91	-9.39	-7.66	-0.84	3.16	5.56	8.07
	Adj. R^2											
Low	0.14	0.28	0.27	0.42	0.51	0.40						
2	0.62	0.60	0.66	0.66	0.56	0.32						
3	0.85	0.86	0.82	0.82	0.85	0.53						
4	0.85	0.93	0.94	0.87	0.87	0.54						
High	0.88	0.94	0.93	0.86	0.65	0.37						

Table 3.4: Double–sorting: Announcement and Non–Announcement Days

This table reports the average daily excess returns (in percentage) of 25 β_{MKT} x β_S equally weighted portfolios by type of day. Portfolios are constructed in the following way: First, stocks are assigned into five quintiles according to their β_{MKT} from lowest (quintile 1) to highest (quintile 5). Then, the stocks within each of the five β_{MKT} quintiles are further sorted into new quintile portfolios based on their monetary policy sensitivity β_S. The row "H–L" refers to the difference in daily returns between the two extreme β_S portfolios. Average returns for 25 portfolios are reported seperately for announcement days in Panel A, non–announcement days in Panel B, and the differences in average returns in Panel C. The right blocks report heteroskedasticity robust t-statistics. Sample period covers the period from January 1, 1995 to December 31, 2009.

		$\boldsymbol{\beta_S}$ **quintiles**											
	$\boldsymbol{\beta_{MKT}}$ **quintiles**	Low	2	3	4	High	H–L	Low	2	3	4	High	H–L
PanelA: A–Day													
		Means						**t–statistics**					
	Low	0.50	0.37	0.35	0.16	0.13	-0.36	4.36	3.11	3.13	1.68	1.67	-3.48
	2	0.58	0.53	0.23	0.26	0.25	-0.34	4.53	4.35	2.44	2.30	2.46	-3.35
	3	0.42	0.58	0.52	0.39	0.27	-0.15	2.95	3.57	3.30	2.92	2.14	-1.52
	4	0.83	0.62	0.55	0.54	0.39	-0.44	4.30	3.59	3.16	3.23	2.81	-3.56
	High	0.93	0.93	0.79	0.64	0.47	-0.46	3.74	4.34	3.73	3.24	2.36	-2.32
PanelB: N–ADay													
		Means						**t–statistics**					
	Low	0.03	0.04	0.04	0.03	0.02	-0.01	1.47	1.77	1.67	1.80	1.41	-0.47
	2	0.03	0.03	0.03	0.03	0.04	0.01	1.19	1.22	1.57	1.37	1.95	0.74
	3	0.04	0.04	0.03	0.03	0.04	0.01	1.64	1.51	1.30	1.15	2.09	0.37
	4	0.04	0.04	0.04	0.06	0.03	-0.01	1.29	1.39	1.52	2.39	1.31	-0.28
	High	0.06	0.05	0.05	0.02	0.04	-0.02	1.59	1.64	1.52	0.80	1.73	-0.71

Continues on next page,

Table 3.5: Double–sorting (continued): Differences

This table reports the average daily excess returns (in percentage) of 25 β_{MKT} x β_S equally weighted portfolios by type of day. Portfolios are constructed in the following way: First, stocks are assigned into five quintiles according to their β_{MKT} from lowest (quintile 1) to highest (quintile 5). Then, the stocks within each of the five β_{MKT} quintiles are further sorted into new quintile portfolios based on their monetary policy sensitivity β_S. The row "H–L" refers to the difference in daily returns between the two extreme β_S portfolios. Average returns for 25 portfolios are reported seperately for announcement days in Panel A, non–announcement days in Panel B, and the differences in average returns in Panel C. The right blocks report heteroskedasticity robust t-statistics. Sample period covers the period from January 1, 1995 to December 31, 2009.

		$\boldsymbol{B_S}$ **quintiles**											
	$\boldsymbol{\beta_M}$ **quintiles**	Low	2	3	4	High	H–L	Low	2	3	4	High	H–L
Panel C: Difference													
		Means						**t–statistics**					
	Low	0.47	0.33	0.31	0.13	0.11	-0.36	4.03	2.71	2.76	1.33	1.36	-3.37
	2	0.55	0.51	0.21	0.23	0.21	-0.35	4.25	4.08	2.11	2.04	2.02	-3.43
	3	0.38	0.55	0.49	0.37	0.22	-0.16	2.65	3.30	3.05	2.68	1.77	-1.56
	4	0.79	0.58	0.51	0.47	0.36	-0.43	4.06	3.34	2.89	2.81	2.55	-3.48
	High	0.87	0.88	0.74	0.61	0.43	-0.44	3.45	4.06	3.47	3.09	2.15	-2.19

continued from previous page.

Table 3.6: Price of Monetary Policy Risk

This table presents the Fama–Macbeth (1973) estimates of prices of risk for (3.2.1). Specification (3.2.1) is estimated on FOMC announcement days with three set of test assets. First column presents the results with 25 β_{MKT} x β_S portfolios as test assets, second column with all individual stocks as test assets and the third column with individual assets and 25 portfolios together as test assets. Robust t–statistics is reported in brackets. The sample period is from January 1, 1995 to December 31, 2009.

	25 Portfolios	**Individual Assets**	**Individual Assets 25 Portfolios**
C	0.160	0.159	0.159
	[1.54]	[1.59]	[1.59]
β_{MKT}	0.424	0.425	0.425
	[2.71]	[2.81]	[2.80]
β_S	-0.066	-0.067	-0.067
	[-3.71]	[-3.69]	[-3.70]
Avg. Adj. R^2	0.32	0.07	0.07

Table 3.7: Prices of Risk for Intraday Windows

This table presents the Fama-Macbeth (1973) estimates of risk prices at the intraday level for three different windows: announcement windows (AW), pre–announcement windows (PAW), post–announcement windows (POAW). Announcement window covers a 30 min period (10 min before to 20 min after). Test assets are the individual stocks on S& P 500. First columns presents the risk price estimates for different windows and the second column reports the related robust t-statistics. Sample period covers the period from January 1, 1995 to December 31, 2009.

AW	est.	t-stat.
C	0.027	1.238
λ_M	-0.111	-1.999
λ_S	0.005	0.863
POAW		
C	0.058	1.470
λ_M	0.036	0.471
λ_S	-0.004	-0.554
PAW		
C	0.010	0.154
λ_M	0.383	4.140
λ_S	-0.033	-2.924

Appendix

3.A Proofs

3.A.1 Precision in the Estimates of Risk Exposures: A Sketch

Let y_t^M and y_t^i respectively denote the logarithm of the instantaneous prices at time t of the market and of each individual stock for $i = 1, 2, \ldots, N$. Assume that the logarithm of the instantaneous price of the market, y_t^M, is driven by the following diffusion process

$$\mathrm{d}y_t^M = \mu^M \mathrm{d}t + \sigma^M \mathrm{d}W_t^M \tag{3.A.1}$$

and assume that dynamics of the logarithm of the instantenous price of each individual stock, y_t^i, are described by the following jump diffusion process,

$$\mathrm{d}y_t^i = \mu^i \mathrm{d}t + \beta_M^i \sigma^M \mathrm{d}W_t^M + \beta_F^i \mathrm{d}J_t^F + \sigma^i \mathrm{d}W_t^i, \tag{3.A.2}$$

where μ^M and μ^i are the drift parameters, W_t^M and W_t^i are standard Brownian motions representing market and idiosyncratic firm–level return innovation. W_t^i and W_t^M are assumed to be orthogonal to each other. $\mathrm{d}J_{F,t}$ is a jump process representing monetary policy shocks with deterministic counting process N_t and jump size distribution being i.i.d $N(\mu_F, \sigma_F^2)$. Exposures to market shocks and to factor jumps are represented by constant β_M^i and constant β_F^i respectively.

Let $r_{t-h,t+H}^i \equiv \int_{t-h}^{t+H} \mathrm{d}y_t^i$ and $r_{t-h,t+H}^M \equiv \int_{t-h}^{t+H} \mathrm{d}y_t^M$, then aggregating over the time intervals $[\tau - h,\ \tau + H]$ surrounding the FOMC announcements (monetary policy shocks) occurring at time τ yields

$$r^i_{\tau-h,\tau+H} \;=\; \mu^i[H-h] + \beta^i_M \int_{\tau-h}^{\tau+H} \sigma^M \mathrm{d}W^M_t + \beta^i_F \sum_{\tau-h\le s\le\tau+H} \Delta J^F_s + \int_{\tau-h}^{\tau+H} \sigma^i \mathrm{d}W^i_t,$$

where $\Delta J^F_s = J^F_s - J^F_{s-}$. Moreover, note that given narrow enough intervals, i.e. windows surrounding the announcements $[\tau-h, \tau+H]$, $\sum_{\tau-h\le s\le\tau+H} \Delta J_{F,s} = J^F_\tau - J^F_{\tau^-} = S_\tau$, where S_τ is distributed with $N(\mu_F, \sigma^2_F)$. Therefore, the above equality can be written as

$$r^i_{\tau-h,\tau+H} = \alpha^i[H-h] + \beta^i_M r^M_{\tau-h,\tau+H} + \beta^i_F S_\tau + \varepsilon_{\tau-h,\tau+H}, \tag{3.A.3}$$

where $\alpha^i \equiv \mu^i - \beta^i_M \mu_M$ and $\varepsilon_{t-h,t+H} \equiv \int_{t-h}^{t+H} \sigma^i \mathrm{d}W^i_t$.

In this setting, the estimates of betas can simply be obtained by regressing the announcement window returns of stock i on the announcement window return of the market and the monetary policy shock measures within the event window and the precision of the estimates of exposure, β^i_F, is given by

$$Var(\hat{\beta}^i_F) = \frac{{\sigma^i}^2[H-h]}{\sigma^2_F}\frac{1}{T}, \tag{3.A.4}$$

where T denotes the number of observations in the sample.

Equation (3.A.4) shows that the sample variance of the exposures to Monetary policy shocks β^i_F, is linearly related to the length of event window with a positive slope. This means that larger event windows leads to higher variances of the estimates of the exposures β^i_F. Accordingly, t–statistics of the estimates declines and R^2 s increases.

Therefore, using returns on individual stocks in narrow windows (i.e. 30 minutes) surrounding the announcements gives us the advantage of obtaining more precise estimates of monetary policy exposures.

3.B Data

3.B.1 Automated High Frequency Data Cleaning Procedure

We are working with trade prices instead of quotes throughout. Barndorff–Nielsen et al. (2009), henceforth BNHLS note that after applying their recommended cleaning procedure, their estimates based on either trade or quote data share a 'remarkable level of agreement' in the context of estimating the quadratic variation of a price process. Our conjecture is their finding should apply when the context is much simpler such as when the interest lies in the plain realized return over a fixed horizon. This naturally extends to the use of trades instead of quotes, and using trades only saves us from reconciling trades with quotes which relies on strong assumptions such as e.g. a Lee–Ready algorithm applying at the five–, one–, or zero–second lag.

Our cleaning procedure consists of the following removing and assigning steps: We only keep trades with sale condition being either blank or of letter code type E, F, @E, @F, @, or T. We proceed by removing all records subject to one of the following three criteria: Occurrence outside the NYSE official cash market trading hours between 9:30 EST and 16:00 EST, zero or negative price entry, zero or negative volume entry. We further filter out corrected trades with TAQ correction indicator other than zero.

The assigning step then allocates prices to the equidistant one second time grid containing 23,400 grid points (6.5 hours $\times$ 60 minutes $\times$ 60 seconds). In case there is more than one trade record sharing the same time stamp (a particular second on a random day), we calculate the volume–weighted average price within that second. BNHLS advocate to use the median price per second whenever multiple transactions per second occur. However, as we use transaction data across all exchanges and the trade size on e.g. BATS is only a fraction of the trade size on NYSE, we believe that the volume–weighted average is a better choice than using the median price whenever transaction data from different exchanges, such as open outcry and fully automated trading, are combined. Without being explicitly mentioned, it is probably the same logic underlying the data cleaning procedure in Bollerslev, Li, Todorov (2015). If there is no transaction recorded during a particular second, we keep the entry from the nearest previous second, denoted forward filtering. In case no transaction has yet been recorded for that day, such as at market opening, we use the nearest subsequent entry, denoted backward–filtering. The backward–filtering has no impact on the calculation of e.g. overnight returns as we supplement the high frequency returns with daily information from the CRSP database.

3.B.2 Cross–Sectional Data Selection: S&P 500 Constituents

The CRSP database provides an accurate list of historical S&P 500 constituents. It is a file containing the CRSP proprietary permanent security identifier PERMNO along with the inclusion and exclusion date. Based on the PERMNOs identified in our sample period, we retrieve the corresponding historical eight digit CUSIPs, denoted NCUSIPs, not to be confused with the permanent CUSIPs in CRSP. These NCUSIPs need to be matched with the twelve digit TAQ CUSIP in order to obtain a list of associated TAQ tickers (trading symbols).[23]

We proceed by deleting the last four digits from the TAQ CUSIP, which correspond to the exchange on which the security is traded, and match the CRSP NCUSIP from which we retrieve a list of trading symbols associated with the NCUSIPs (TAQ CUSIPs). We then assign the latest updated ticker symbol to the CRSP NCUSIP

[23] Note that the CRSP ticker is not necessarily equal to the TAQ ticker and thus the TAQ trading symbol. CRSP provides a trading symbol (TSYMBOL) but the field is often blank and thus of limited use.

and thus to the CRSP PERMNO.

Bibliography

Andersen, T.G., and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review* 39, 885–905. 14, 64

Andersen, T.G., T. Bollerslev, and S. Lange (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* 6, 457–477. 76

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, 42–55. 64

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579–625. 14, 15, 17, 64

Andersen, T.G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macroeconomic announcements: Real–time price discovery in foreign exchange. *American Economic Review* 93, 38–62. 118, 125

Andersen, T.G., , T. Bollerslev, F. X. Diebold, and C. Vega (2007). Real–time price discovery in global stock, bond, and FX markets. *Journal of International Economics* 73, 251–77. 125

Andersen, T.G., , T. Bollerslev, and F. X. Diebold (2007). Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics & Statistics* 89, 701–720. 14

Andersen, T.G., T. Bollerslev, and N. Meddahi (2004). Analytic evaluation of volatility forecasts. *International Economic Review* 45, 1079—1110. 14

Andersen, T.G., D. Dobrev, and E. Schaumburg (2014). A robust neighborhood truncation approach to estimation of integrated quarticity. *Econometric Theory* 30, 3—59. 15

Ang, A., J. Liu, and K. Schwarz (2010) Using stocks or portfolios in tests of factor models. *Working Paper.* 132

Balduzzi, P. and C. Robotti (2008). Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business & Economic Statistics* 26, 354–368. 25, 29

Bandi, F. M. and J. R. Russell (2006). Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655–692. 15

Barndorff–Nielsen, O.E., P.E. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal* 12, 1–32. 15, 127, 147

Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 167–241. 14, 64

Barndorff-Nielsen, O.E. and N. Shephard (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 253–280. 14, 64, 84

Barndorff-Nielsen, O.E. and N. Shephard (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17, 457–477. 14, 64

Barndorff-Nielsen, O.E. and N. Shephard (2003). Realized Power Variation and Stochastic Volatility Models. *Bernoulli* 9, 243–265. 64

Barndorff-Nielsen, O.E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–37. 64

Bernanke, B.S. and K. N. Kuttner (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of Finance* 60, 1221–1257. 17, 118, 120, 125, 126, 135

Best, Michael J and Grauer, Robert R (1991). On the sensitivity of mean–variance efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies* 4, 315–342. 22

Black, Fischer (1993). Estimating expected return. *Financial Analysts Journal* 36–38. 20

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica* 77, 623–685. 118

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta–Eksten, and S.J. Terry (2014). Really uncertain business cycles. *Working Paper.* 118

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore. 79

Bollerslev, T., S.Z. Li, and V. Todorov (2015). Roughing up beta: Continuous vs. discontinuous betas, and the cross–section of expected stock returns. *Working Paper.* 148

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer-Verlag, Berlin.

Douglas T. Breeden (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7, 265–296. 20, 24

Broadie, Mark (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research: Special Issue on Financial Engineering* 45, 21–58.

Brown, L.D., and M.G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics* 6, 2384–2398. 103

Buchanan, H.E. and T.H. Hildebrandt (1908). Note on the Convergence of a Sequence of Functions of a Certain Type. *Annals of Mathematics* 9, 123–126. 74

Christensen, K., R.C.A Oomen, and M. Podolskij(2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114, 3, 576–599. 15

Chopra, V.K. and , W.T. Ziemba (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19, 6–11. 22

Christiano, L.J., R. Motto, and M. Rostagno (forthcoming). Financial factors in economic fluctuations. *American Economic Review*. Forthcoming. 118

Cieslak, A., Morse A., and Vissing–Jorgensen, A. (2015). Stock returns over the FOMC cycle. *Working Paper* 121, 123

De Goeij, P. C., Hu, J., and Werker, B. J. M (2013). The price of macroeconomic news. *Working Paper* 120

Clément, E., S. Delattre, and A. Gloter (2013). An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications* 123, 2500–2521. 64, 65, 81, 93, 94

Cochrane, J.H. (2001). *Asset pricing.* New Jersey: Princeton University Press. 23, 27

Comte, F., and E. Renault (1998). Long-Memory in Continuous Time Stochastic Volatility Models. *Mathematical Finance* 8, 291-323. Reprinted in Stochastic Volatility: Selected Readings, N. Shephard (ed.), Oxford University Press, 2005. 14, 65

Davidson, J. (2002). *Stochastic limit theory.* 2nd edition. Oxford University Press, New York.

DeMiguel, V., L. Garlappi, , F.J. Nogalesand, and R.Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55, 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22, 5, 1915–1953. 22, 46

Fama, E.F. (1998). Determining the number of priced state variables in the ICAPM. *Journal of Financial and Quantitative Analysis*, 33. 25

Fama, E.F. and K.R. French (1992). The cross–section of expected stock returns. *The Journal of Finance* 47,2 427–465. 30, 31, 44

Fama, E.F. and K.R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56. 21, 30, 31, 39

Fama, E.F. and J.D. MacBeth (1973). Risk, return, and equilibrium: Empirical Tests. *The Journal of Political Economy* 81, 607–636.

Fatum, R. and B. Scholnick (2008). Monetary policy news and exchange rate responses: Do only surprises matter? *Journal of Banking & Finance* 32, 1076–1086. 124

Fleming, M.J. and Piazzesi, M. (2005). Monetary policy tick–by–tick. *Working Paper.* 125

Ferson, W.E. and C.R. Harvey (1993). The risk and predictability of international equity returns. *The Review of Financial Studies* 6, 527–566.

Frost, P.A. and J.E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial Quantitative Analysis*, 21, 293–305. 22

Frost, P.A. and J.E. Savarino (1988). For better performance constrain portfolio weights. *Journal of Portfolio Management* 15, 29–34. 22

Garcia, R. and E. Renault (1998). A note on hedging in ARCH and stochastic volatility option pricing models. *Mathematical Finance* 2, 153–161.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 445—475. 14

Gilchrist, S., J. Sim, and E. Zakrajsek (2009). Uncertainty, credit spreads and investment dynamics. *Working Paper.* 118

Gorodnichenko, Y. and M. Weber (2014). Are sticky prices costly? Evidence from the stock market. *Working Paper.* 125

Gürkaynak, R. S., B.P. Sack, and E.T. Swanson (2005). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *International Journal of Central Banking* 1, 55–94. 17, 118, 120, 122, 125

Gürkaynak, R.S. and J. Wright (2013). Identification and inference using event studies. *Technical report.* 122

Hall, A.R. (2005). *Generalized method of moments, Advanced Texts in Econometrics.* Oxford: Oxford University Press. 52

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 1029–1054. 26

Hautsch, N., L. M. Kyj, and P. Malec (2015). Do high–frequency data improve high–dimensional portfolio allocations? *Journal of Applied Econometrics* 30,2 263–290. 15

Hayashi, T., J. Jacod, and N. Yoshida (2011). Irregular sampling and central limit theorems for power variations: The continuous case. *Annales de l'Institut Henri Poincaré* 47, 1197–1218. 66, 70, 84, 87

Hou, K. and R.L. Kimmel (2010). On estimation of risk premia in linear factor models. *Working Paper*. 25, 28

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3,4 456—99.

Huberman, G., S. Kandel, and R.F. Stambaugh (1987). Mimicking portfolios and exact arbitrage pricing. *The Journal of Finance* 42, 1–9. 25

Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta pricing models using cross–sectional regression. *The Journal of Finance* 53, 1285–1309. 30

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58, 4 1651–1684. 46

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications* 118, 517–559. 64, 78

Jacod, J. and M. Rosenbaum (2013). Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics* 41, 1462–1484. 15, 16, 63, 64, 65, 66, 67, 68, 82, 83, 84, 85, 89, 90, 94

Jacod, J., and A.N. Shiryaev (2002). *Limit Theorems for Stochastic Processes.* 2nd edition. Springer-Verlag, Berlin. 106

Jobson, J. D and B. Korkie (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association* 544–554. 22

Jobson, J. D and B. Korkie (1981). Putting Markowitz theory to work. *Journal of Portfolio Management* 7, 70–74. 22

Jorion, P. (1991). Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking and Finance* 717–727. 20, 46

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two–pass cross–sectional regression methodology. *The Journal of Finance* 68, 6, 2617–2649. 30

Kan, R. and C. Robotti (2011). On the estimation of asset pricing models using univariate betas. *Economics Letters* 110, 117–121. 30

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two–pass cross–sectional regression methodology. *The Journal of Finance* 68, 6, 2617–26. 30

Kandel, S. and R.F. Stambaugh (1995). Portfolio Inefficiency and the Cross-Section of Expected Returns. *The Journal of Finance*, 50, 1, 157–184. 25

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, 149, 2, 149–173. 30

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 26, 60–93. 63, 64, 67, 84, 86, 87

Kuttner, K.N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics* 47, 523–44. 17, 118, 120, 124, 126

Lamont, O.A. (2001). Economic tracking portfolios. *Journal of Econometrics*, 105, 161–184. 25

Le Cam, L.L. (1986). *Asymptotic methods in statistical decision theory.* Springer-Verlag, Berlin. 101

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A cross–sectional test when risk premia are time–varying. *Journal of Political Economy* 109, 1238–1287. 20, 21, 24

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96, 175–194. 30

Li, Q., M. Vassalou, and Y. Xing(2006). Sector investment growth rates and the cross section of equity returns. *The Journal of Business* 79, 1637–1665. 21

Li, Y., P.A. Mykland, E. Renault, L. Zhang, and X. Zheng (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory* 30, 580–605. 70

Li, J., G. Tauchen, and V. Todorov (2013). Volatility occupation times. *The Annals of Statistics* 41, 1865–1891. 78

Litterman, R. (2003). *Modern investment management: An equilibrium approach.* New Jersey: John Wiley & Sons. 22

Lucca, M. and E. Moench (2015). The pre-FOMC announcement drift. *The Journal of Finance* 70, 329–371. 17, 18, 123, 128, 135

Lustig, H.N. and S.G. Van Nieuwerburgh (2005). Housing collateral, consumption insurance, and risk premia: An empirical perspective. *The Journal of Finance* 60, 1167–1219. 21

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* 7, 1, 77—91. 13, 22, 42

Maheu, J. M. and T. H. McCurdy (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160,1 69–76. 15

Meyer, B. (1984). Some Inequalities for Elementary Mean Values. *Mathematics of Computation* 42, 193–194.

Merton, R C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41, 867-887. 20, 24, 135

Merton, R.C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361. 22, 76

Michaud, R.O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal* 45, 1, 31—42. 22

Mykland, P.A., and L. Zhang (2006). ANOVA for diffusions and Itô processes. *The Annals of Statistics* 34, 1931–1963. 68, 74, 75

Mykland, P.A., and L. Zhang (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–1445. 63, 64, 67, 68, 81, 82, 84, 85, 94

Mykland, P.A., and L. Zhang (2012). The econometrics of high frequency data. In M. Kessler, A. Lindner, and M. Sørensen (eds.), *Statistical methods for stochastic differential equations*, pp. 109-190. CRC Press - Chapmann and Hall 75, 85

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *The Annals of Statistics* 24, 2399–2430. 102

Protter, P. (1995). *Stochastic integration and differential equations: A new approach.* Springer-Verlag, Berlin. 95, 96

Reiß, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics* 39, 772–802. 65, 72, 93, 98

Renault, E. (1997). Econometric models of option pricing errors. In D.M. Kreps and K.F. Wallis (eds.), *Advances in economics and econometrics: Theory and applications*, pp 223–278. Cambridge University Press.

Renault, E., Th. van der Heijden, and B.J.M. Werker (2014). The dynamic mixed hitting–time model for multiple transaction prices and times. *Journal of Econometrics* 180, 233–250.

Revuz, D. and M. Yor (1991). *Continuous martingales and brownian motion.* Springer-Verlag, Berlin. 96

Rigobon, R. and B. Sack (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics* 51, 1553–1575. 17, 118, 122

Romer, D.H. and C.D. Romer (2000). Federal reserve information and the behavior of interest rates. *American Economic Review* 90, 429–457. 120

Ross, S.A. (1976a). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341 - 360. 20, 118

Ross, S.A. (1976b). Return, risk, and arbitrage. *Wharton Discussion Paperr, 1973, published in I. Friend and J. Bicksler, eds., Risk and Return in Finance. (Cambridge: Ballinger)* 1976, 189-217. 20

Savor, P. and M. Wilson (2013). How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis* 48, 343–375. 17, 120, 121

Santos, T. and P. Veronesi (2006). Labor income and predictable stock returns. *The Review of Financial Studies* 19, 1–44. 21

Shanken, J. (1992). On the estimation of beta–pricing models. *The Review of Financial Studies* 5, 1, 1–55. 30

Shanken, J. and G. Zhou (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics* 84, 1, 40–86. 30

Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19, 425–442. 20

Stock, J.H. and M.W. Watson (2012). Disentangling the channels of the 2007–2009 recession. *Working Paper.* 118

Taylor, S.J. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317–340. 76

Todorov, V. and G. Tauchen (2012). The realized laplace transform of volatility. *Econometrica* 80, 1105–1127. 15, 78, 83

Van der Vaart, A.W. (2000). *Asymptotic statistics.* Cambridge University Press, Cambridge. 75, 76, 78, 113

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235–250. 66, 90